

Lehigh University

Lehigh Preserve

---

Theses and Dissertations

---

2019

## Parallelism, constraint, and functional genome evolution in experimentally evolving populations of *Saccharomyces cerevisiae*

Kaitlin Fisher

Lehigh University, [kjf214@lehigh.edu](mailto:kjf214@lehigh.edu)

Follow this and additional works at: <https://preserve.lehigh.edu/etd>



Part of the [Integrative Biology Commons](#)

---

### Recommended Citation

Fisher, Kaitlin, "Parallelism, constraint, and functional genome evolution in experimentally evolving populations of *Saccharomyces cerevisiae*" (2019). *Theses and Dissertations*. 5594.

<https://preserve.lehigh.edu/etd/5594>

This Dissertation is brought to you for free and open access by Lehigh Preserve. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Lehigh Preserve. For more information, please contact [preserve@lehigh.edu](mailto:preserve@lehigh.edu).

Parallelism, constraint, and functional genome evolution in experimentally evolving  
populations of *Saccharomyces cerevisiae*

by

Kaitlin J Fisher

A Dissertation

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Biological Sciences

Lehigh University

May, 2019

© 2019 Copyright  
Kaitlin J Fisher

Approved and recommended for acceptance as a dissertation in partial fulfillment  
of the requirements for the degree of Doctor of Philosophy

Kaitlin J Fisher  
Parallelism, constraint, and functional genome evolution in experimentally evolving  
populations of *Saccharomyces cerevisiae*

April 26, 2019

---

Defense Date

April 26, 2019

---

Approved Date

---

Dissertation Director  
Greg Lang, PhD

Committee Members

---

Amber Rice, PhD

---

Michael Kuchka, PhD

---

Molly Burke, PhD

## Acknowledgements

I first would like to acknowledge my mentor, Greg Lang. I was one of the first graduate students to join Greg's lab, and I did so principally because I was drawn to the research system he intended to build his lab around. This turned out to be a serendipitous choice and, research system aside, I could not have asked for a better advisor. Greg provided me with structure and guidance while at the same time affording me a long intellectual leash with room to make - and learn from - mistakes. It would be difficult to quantify how much I learned from Greg and I am immensely grateful for all that he has invested in me. I am the first in what I'm sure will be a long list of trainees that will profit from the care that Greg puts into mentorship.

I would like to acknowledge that much of the work described in this dissertation came about through collaborations. At points throughout my time at Lehigh I worked closely with members of my own lab, Sean Buskirk and Ryan Vignogna, as well as extramural collaborator Sergey Kryazhimskiy. I learned an immense amount from these individuals and profited from the supplemental mentorship they provided. I've found the most valuable parts of these collaborations to be those instances where there has been disagreement. I want to thank those that I've worked closely with for the criticisms and objections that, ultimately, improved the quality of the work we put out.

I also want to broadly acknowledge my lab-mates beyond their roles as collaborators and coauthors. Sean Buskirk, Dimitra Aggelli, Ryan Vignogna, and Dan Marad have all been wonderful colleagues and friends during my time here. My research has benefited immensely from the collaborative environment of our lab and the constant flow of ideas.

The small size of the department here at Lehigh has meant that I have gotten to know many of the faculty, research staff, and graduate students well. I have benefited from great conversations with many of these colleagues. I am particularly grateful to Layla Al-Shaer, who has been a great friend and a great sounding board.

I would like to thank my committee members, Amber Rice, Michael Kuchka and Molly Burke, for their guidance, helpful feedback, and the enthusiasm they've shown for my work. My research direction took several turns during my time at Lehigh and I want to thank my committee members for guiding me to accept the reality that sometimes-experimental plans simply don't pan out.

My family has not always understood the academic path I'm on or why I chose it, but they have always remained supportive. I want to thank my sisters for their encouragement and I want to thank my parents for instilling a work ethic in me that has helped me navigate the demands of graduate school.

Finally, I would like to thank my best friend and partner, Jason. The pursuit of a graduate degree is accompanied certain sacrifices, and Jason has unflinchingly taken in stride all of the hardships that my insistence on pursuing science has produced. He has been unwavering in his support and he has been my biggest cheerleader.

## Table of Contents

<b>List of Figures</b>	<i>vii</i>
<b>List of Tables</b>	<i>viii</i>
<b>Abstract</b>	<b>1</b>
<b>Chapter 1.</b> Introduction	<b>2</b>
<b>Chapter 2.</b> Parallelism of adaptive genome duplication	<b>12</b>
<b>Chapter 3.</b> Constraints imposed by dominance underlie parallelism in experimental evolution	<b>53</b>
<b>Chapter 4.</b> Leveraging parallelism to detect genetic interactions between evolved mutations.	<b>73</b>
<b>Appendix A.</b> Statistical method for detection of genes harboring beneficial mutations	<b>104</b>
<b>Appendix B.</b> Supporting material for chapter 2	<b>111</b>
<b>Appendix C.</b> Preliminary findings on the evolutionary dynamics of gene-drives	<b>131</b>
<b>References</b>	<b>131</b>
<b>Vita</b>	<b>146</b>

## List of Figures

<b>Chapter 2</b>		
2-1	Loci of evolved mutations	18
2-2	Fitness effects and population dynamics of autodiploids	20
2-3	Time-course ploidy dynamics of 16 focal populations	22
2-4	Time-course ploidy and fitness dynamics of 13 focal populations	23
2-5	Forward simulations of ancestral autodiploid dynamics	25
2-6	Common targets of selection	27
2-7	Loci of evolved homozygous mutations	32
2-8	Evolved aneuploidies	33
2-9	Spore viability of evolved clones	36
2-10	Validation of hydroxyurea arrest assay	48
<b>Chapter 3</b>		
3-1	Fitness effects of evolved <i>STE4</i> mutations and deletions in diploids	62
3-2	Fitness effects of evolved <i>STE4</i> mutations in haploids	64
3-3	Fitness of synonymous mutation introduced in strain construction	65
3-4	Phylogenetic evidence of overdominance at <i>STE4</i>	67
3-5	Evolved <i>WHI2</i> allele fitness and rates of LOH	68
<b>Chapter 4</b>		
4-1	Histograms of observed and simulated $MI_{ij}$ values	85
4-2	Simulated null distribution of $MI_{tot}$ values	86
4-3	Effect of pseudocount variation on $MI_{tot}$	87
4-4	Example individual gene-pair $MI_{ij}$ simulations	88
4-5	False discovery rate estimates	90
4-6	Phylogeny of all populations with a <i>TRK1</i> and/or <i>PHO84</i> mutation	94
4-7	<i>TRK1/PHO84</i> double mutant fitness assay	96
4-8	Modular structure of significant gene pairs	97
4-9	Hierarchical clustering of observed $MI_{ij}$ values	98
<b>Appendix A</b>		
A-1	Common genic targets of selection across three experiments	107
<b>Appendix C</b>		
C-1	A synthetic gene drive allele at <i>ADE2</i>	121
C-2	A synthetic gene drive allele at <i>HIS3</i>	123
C-3	Fluorescence approach to quantifying genotype frequencies	124
C-4	Dynamics from short-term <i>ADE2</i> drive experiment	127



## List of Tables

<b>Chapter 2</b>		
2-1	Autodiploidization reports in past literature	14
2-2	Evolved mutations identified in common targets of selection	29
2-3	Copy number polymorphisms in evolved clones	35
<b>Chapter 3</b>		
3-1	Evolved <i>STE4</i> mutations	63
<b>Chapter 4</b>		
4-1	Identities of 33 significant pairwise interactions	91
4-2	Evolved alleles of genes in top 5 putative interactions	92
<b>Appendix B</b>		
B-1	Strains generated for <i>STE4</i> assays	114

# Abstract

From sticklebacks to insects to mice, closely related populations encountering similar selective pressures tend to evolve parallel phenotypes. The era of genomics has revealed that some instances of parallel phenotypes are caused by similar mutations in similar genes. Others cases of parallel adaptation are shown to be driven by mutations in distinct genes. These observations spark an open question in evolutionary biology; given identical starting points and selective pressures, how repeatable should we expect molecular adaptation to be? Laboratory evolution of model organisms presents an ideally suitable system for beginning to answer this question. This work uses experimental evolution of *Saccharomyces cerevisiae* combined with genomics and functional genetics to ask questions about the repeatability of evolution. We focus on identifying and examining the adaptive consequence of molecular events that arise more often than expected by chance. We find extensive parallelism in ploidy evolution when genome duplication proves adaptive. We use whole genome sequencing to identify loci targeted by selection multiple independent times across populations. We use one of these loci, *STE4*, to examine how dominance constrains mutation and adaptation. We then leverage the extensive gene-level parallelism we observe to detect genetic interactions and measure the effect of epistasis on genotype evolution.

# Chapter 1

## Introduction

*Note- This chapter contains some material drawn from a review paper<sup>1</sup> coauthored with my advisor, Greg Lang.*

A single historical timeline has produced all of the diversity of life we observe today. Evolutionary biologists have often inferred evolutionary histories of populations based on observations, and hypotheses are generated to explain these inferred histories and provide mechanistic insight into the evolutionary process. Yet testing specific hypotheses regarding mechanisms of evolution is difficult because we cannot say anything about the countless other possible but unrealized evolutionary histories. In his 1989 book, *Wonderful Life*, Stephen Jay Gould proposed the following thought experiment: rewind the tape of life and let evolution play out a second time. In doing so, does the replay produce anything like what we see today? In other words, is evolution reproducible, or would the randomness inherent in evolutionary processes change histories and produce wildly different outcomes? The birth of the modern field of experimental evolution closely followed Gould's book and has allowed the pursuit of this central question to move from philosophical to empirical.

---

<sup>1</sup> Fisher, K. J., & Lang, G. I. (2016). Experimental evolution in fungi: An untapped resource. *Fungal Genetics and Biology*, 94, 88-94.

Another way of asking about the reproducibility of evolution is asking how frequently we observe parallel evolution. Parallel evolution describes the biological phenomena by which two phylogenetically distinct lineages evolve analogous features in response to similar selective pressures. Determining the prevalence of parallelism amongst adapting lineages is not trivial. In some fortuitous situations, natural replication allows one to follow evolution in several independent replicate populations. These “natural experiments” have been well studied, for example in Galapagos finches (Grant, Grant, Markert, Keller, & Petren, 2004), *Astyanax* cavefish (Protas et al., 2006), and sticklebacks (Hohenlohe et al., 2010; Jones et al., 2012).

### **Parallelism in “natural replicates”**

Synthesis of “natural replicate” findings reveals a strong signature of parallelism in separately evolving populations facing the same selective pressures. Phenotypic parallelism is exceedingly common amongst sister taxa that have repeatedly invaded similar but separate habitats. For example, distinct populations of the Mexican cave tetra (genus *Astyanax*) that have colonized cave habitats have independently undergone eye and pigmentation loss (Protas et al., 2006; Wilkens & Strecker, 2003). Similarly, freshwater sticklebacks (genus *Gasterosteus*) have repeatedly colonized freshwater habitats and become geographically isolated from ancestral marine populations. A number of morphological differences relative to marine populations are shared among independent freshwater populations, including the well-studied reduced gill rakers (Hagen & Gilbertson, 1972; Moodie & Reimchen, 1976) and pelvic structures (Bell, 1987). Parallel phenotypes can also be seen at the level of gene regulation; parallel changes in gene expression in opsin genes have been documented among closely related but genetically

isolated cichlid (family Cichlidae) populations that are adapting to similar depths (O'quin, Hofmann, Hofmann, & Carleton, 2010).

Phenotypic parallelism suggests that there are only so many solutions to a given problem and that selection will steer towards one of them. Some cases of remarkably parallel phenotypes in “natural replicate” populations are due to causal mutations in distinct genes and pathways. Such is the case with convergent pigmentation patterns in oldfield mice (*Peromyscus polionotus*). Populations of mice living in beach habitats on the Atlantic and Gulf have independently acquired a lighter, more cryptic coloration. Pigmentation adaptation in Gulf populations seems to be mediated by mutations directly to the *Mc1r* gene, while Atlantic populations appear to have adapted via a completely different genetic route (Steiner, Römler, Boettger, Schöneberg, & Hoekstra, 2008).

Commonly, however, phenotypic convergence amongst closely related taxa is attributable to underlying parallelism at the genetic level, such that recurrent phenotypes can be attributed to variation in the same gene or group of functionally related genes. The best examples of this come from the two teleost species mentioned above, the Mexican cave tetra and the freshwater stickleback. In both systems, small populations that have repeatedly become isolated in similar new environments display evidence of parallel molecular routes to adaptation (Chan et al., 2010; Colosimo et al., 2005; Glazer, Cleves, Erickson, Lam, & Miller, 2014; Protas et al., 2006). Molecular convergence is also seen across broader taxonomic groups including waterfowl adapted to high altitude (McCracken et al., 2009), arthropod herbivores that overcome plant defenses (Zhen, Aardema, Medina, Schumer, & Andolfatto, 2012), marine mammals (Foote et al., 2015),

cold-hardy conifer species (Yeaman et al., 2016), and mammals that consume bamboo (Hu et al., 2017), to name just some examples.

### **Experimental evolution is ideally suited to study parallelism**

Extensive parallelism in natural populations does suggest evolution is highly reproducible, but such observations are limited in interpretability. Natural experiments are not perfect replicates. Details of the environments will differ and the number of replicates is constrained. Field studies are also constrained by incomplete characterization of ancestral populations, making it difficult to distinguish between ancestral variation and parallel *de novo* evolution. These limitations of field studies are overcome in laboratory evolution experiments. Experimental evolution complements the study of natural populations and provides a system in which specific hypotheses can be tested. At its core, the field of experimental evolution is the realization of Gould's thought experiment. An evolution experiment involves initiating a few, hundreds, or even thousands, of initially identical populations and passaging them forward through time to assess the full distribution of evolutionary outcomes given a set of initial conditions. The longest running evolution experiment began in 1988, when Richard Lenski initiated 12 replicate cultures of *E. coli* which have been propagated daily for the last 28 years, surpassing 64,000 generations of growth. The Lenski long-term evolution experiment (LTEE) is the iconic example of laboratory experimental evolution.

Exchanging the complexities of the natural environment for the simplicity of the laboratory provides a number of advantages. Most microbes can be cryogenically archived to generate "frozen fossil records" that can be returned to at any time in order to identify mutations or measure fitness. Evolutionary parameters that are difficult to

quantify in natural populations (such as population size and mutation rate) can be precisely measured and controlled. Selection can be tightly controlled by strictly defining media, temperature, and other growth conditions. Genetic variation can be defined at the onset of any experiment and gene flow can be absent or modulated. Population size and bottlenecks can be accurately quantified and kept constant, mitigating the effect of genetic drift. Gains in fitness can be tracked experimentally and high throughput next-generation sequencing can be applied to link changes in phenotype with underlying mutations. The large research communities devoted to the study of the model systems afford a number of tools for the genetic manipulation and the genetic and genomic analyses of evolved populations and provide a meaningful context in which to interpret the effects of individual mutations on fitness and other relevant phenotypes.

Findings from diverse evolution experiments bolster those from “natural replicates” and indicate that, in the face of identical selection, marked phenotypic divergence is rare while phenotypic parallelism is common. Fitness itself characteristically increases in a similar manner across replicate populations (Lenski, 2017), reaches similar peaks (Wiser, Ribeck, & Lenski, 2013), and can remain similar across environments (Bailey, Rodrigue, & Kassen, 2015). Reproducibility of non-fitness trait evolution is evident in the *E. coli* LTEE; parallel increases in cell size (Lenski, & Travisano, 1994), parallel shifts in catabolism (Cooper, Schneider, Blot, & Lenski, 2001), and parallel changes to both gene expression and proteomes (Cooper, Rozen, & Lenski, 2003; Pelosi et al., 2006). Shorter-term studies have also found parallelism in non-fitness traits (Fong, Joyce, & Palsson, 2005; Van Ditmarsch et al., 2013). Replicate populations do occasionally evolve

divergently (Hillesland, Velicer, & Lenski, 2008; Ratcliff et al., 2013), but these cases are rare in the absence of environmental heterogeneity.

Similar to “natural replicate” populations, experimental lineages evolving under identical selective pressures commonly adapt via parallel molecular routes. Studies that have directly addressed genic parallelism find it to be widespread (Betancourt, 2009; Bull et al., 1997; Deatherage, Kepner, Bennett, Lenski, & Barrick, 2017; Lenski, Richard E., 2017; Tenailon et al., 2012). Most studies do not directly quantify parallelism *per se*, but still report results indicative of highly repeatable molecular dynamics. Applying high-throughput sequencing to experimentally evolved populations often identifies the same genes accumulating mutations in replicate populations (Kvitek & Sherlock, 2011; Lang et al., 2013; Venkataram et al., 2016). Copy number variations (CNVs) containing specific genes are reproducible outcomes of nutrient limitation or chemical stress (Adamo, 2012; Gresham et al., 2010; Payen et al., 2014). Finally, parallelism in ploidy changes in eukaryotic microbes is being increasingly reported (as will be addressed in the first chapter of this dissertation) (Gerstein, Chun, Grant, & Otto, 2006; Gorter et al., 2017; Hong & Gresham, 2014).

### **Interplay of parallelism and constraint**

Extensive parallelism may, in part, indicate a pronounced effect of genetic constraint on sequence evolution. Genetic constraint broadly refers to factors that limit the mutational trajectories accessible to an evolving genome. Sources of constraint can include factors that limit the number of mutations that can produce a given trait value, such as low genomic redundancy wherein adaptive traits have few underlying loci (Chevin, 2013). Constraint can also be imposed by factors that affect the fitness effects of



new mutations as well as those that limit the accessibility of beneficial mutations (Connallon & Hall, 2018). Alternatively, constraint can indirectly be imposed by mutational bias; the most frequently occurring mutation type that produces an advantageous phenotype is most likely to be observed (Stoltzfus & McCandlish, 2017; Storz, 2016). Finally, epistatic interactions between genes can constrain sequence evolution of interacting genes (Storz, 2016).

Constraints imposed by low genomic redundancy and those imposed factors affecting the fitness effects of new mutations are difficult to disentangle. The same gene may be frequently mutated to achieve a given trait value because that is the only gene (or one of few) that can produce that trait value. Conversely, many mutations may be able to generate a given phenotype, however these mutations are inaccessible for reasons unrelated to the adaptive trait space. Modeling of trait evolution under both sources of constraint suggests an affect of both in shaping convergence (Yeaman, Sam, Gerstein, Hodgins, & Whitlock, 2018). In the case of microbial experimental evolution, wherein the reproductive mode is asexual clonal expansion, constraint due to accessibility of mutations has a particularly pronounced affect (Gerstein, Kuzmin, & Otto, 2014; Marad, Buskirk, & Lang, 2018). This is largely due to constraints imposed by dominance, which scale the likelihood of alleles fixing in populations independent of their relative fitness effects (Haldane, 1924). Mutations in diploid and polyploid backgrounds should need to exhibit at least partial dominance in order to be fixed by selection, while completely recessive beneficial mutations rely on instances of loss-of-heterozygosity (Gerstein et al., 2014; Smukowski Heil et al., 2017).

Mutational bias is a source of constraint that has been receiving increasing attention in recent years (Stoltzfus & McCandlish, 2017; Storz, 2016). Rather than operating by filtering out possible mutational routes (as would be the case with low genomic redundancy and limited accessibility), mutational bias operates via distorting the relative frequencies of certain adaptive events. Adaptive events that occur at higher rates are more likely to be observed than those that are rare. The best examples of this from both comparative genomics and experimental evolution are structural rearrangements. Amplification and/or deletion events producing copy number variants (CNVs) vary in frequency across the genome, with some loci being prone to frequent copy number fluctuation (Brewer et al., 2015; Press, Hall, Morton, & Queitsch, 2018). CNVs accordingly represent some of the most marked examples of parallel genome evolution in experimental evolution (Fisher, Buskirk, Vignogna, Marad, & Lang, 2018; Payen et al., 2014; Sanchez et al., 2017). Signatures of parallelism in gene amplification and deletion are also documented broadly across diversifying clades (Clop, Vidal, & Amills, 2012; Stratton, Campbell, & Futreal, 2009; Żmieńko, Samelak, Kozłowski, & Figlerowicz, 2014). In the case of pelvic reduction in sticklebacks, the causal recurrent deletions are found in an enhancer element (Chan et al., 2010) that has recently been shown to exhibit increased levels of fragility, leading to high rates of local deletions (Xie et al., 2019).

The final source of constraint discussed here is epistatic interactions. Epistatic constraint can be usefully broken down into two types: intralocus epistasis and interlocus epistasis. Intralocus epistasis between sites within a coding sequence has well demonstrated effects of constraining protein evolution and entrenching evolved mutations (Gong et al., 2013; Bridgham et al., 2009; Lunzer et al., 2010; Shah et al., 2015). A great

deal less is known about how inter locus epistasis – also referred to as genetic interactions – influences adaptive trajectories. Genetic interaction is used to describe the phenomenon by which the phenotype (or fitness) of a double mutant deviates from what is expected given the phenotypes of single mutants (reviewed in Costanzo et al., 2019). Given the dense and interconnected yeast interactome (Costanzo et al. 2019, Baryshnikova et. al., 2016; Costanzo et al., 2016), interactions between mutations in different genes are expected to contribute to, and constrain, adaptation.

## **Summary**

Parallelism and homoplasy used to be viewed as exceptional in evolutionary theory. In a circular manner, this view, and its application in “maximum parsimony” has traditionally shaped phylogenies in such a way that repeated evolution was thought to be rare. The genomics era has revealed that evolution may be more reproducible than previously thought. Groups that used to be considered monophyletic based on morphology have been revealed to be polyphyletic with widespread homoplasy (Chueca, Gómez-Moliner, Madeira, & Pfenninger, 2018; Parra-Olea & Wake, 2001; Wu et al., 2015). In this dissertation, I will directly examine the extent of parallelism of specific molecular events, the sources of constraint that produce parallelism, and the consequences and applications of molecular parallel evolution. In Chapter 1 I focus on parallel ploidy evolution and the consequential constraint imposed by ploidy dynamics on sequence evolution. The extent to which microbes adapt via rapid ploidy fluctuations has become appreciated only recently. My work in Chapter 1 suggests that ploidy change is a fairly certain outcome when the selective benefit is sufficiently high. In Chapter 2 I focus on a single locus that acquires mutations in a specific gene region multiple times across

different experimentally evolved populations. I find that a complex interplay of underdominance and overdominance explains the observed patterns of parallelism. To my knowledge this is the first empirical demonstration of how dominance affects sequence evolution in real time. Finally, in chapter 3 I leverage parallelism as a tool to examine the pervasiveness of epistasis between experimentally evolved mutations. This computational investigation revealed that parallelism can be used to measure signatures of epistasis, as well as identify interactions between specific gene pairs.

## Chapter 2

### Parallelism of adaptive genome duplication

*Note - The work described in this chapter has been published<sup>1</sup> in collaboration with Sean Buskirk. As co-first author, Sean performed most of the bioinformatic analyses for this paper. Ryan Vignogna performed the simulation assays described in this chapter. The datasets referenced in this chapter have been archived with the publication<sup>2</sup>.*

#### **Abstract**

Genome duplications are important evolutionary events that impact the rate and spectrum of beneficial mutations and thus the rate of adaptation. Laboratory evolution experiments initiated with haploid *Saccharomyces cerevisiae* cultures repeatedly experience whole-genome duplication (WGD). We report recurrent genome duplication in 46 haploid yeast populations evolved for 4,000 generations. We find that WGD confers a fitness advantage, and this immediate fitness gain is accompanied by a shift in genomic and phenotypic evolution. The presence of ploidy-enriched targets of selection and structural variants reveals that autodiploids utilize adaptive paths inaccessible to haploids. We find that autodiploids accumulate recessive deleterious mutations, indicating an increased susceptibility for nonadaptive evolution. Finally, we report that

---

<sup>1</sup>Fisher, K. J., Buskirk, S. W., Vignogna, R. C., Marad, D. A., & Lang, G. I. (2018). Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*. *PLoS genetics*, 14(5), e1007396.

<sup>2</sup><https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007396#sec023>

WGD results in a reduced adaptation rate, indicating a trade-off between immediate fitness gains and long term adaptability.

## **Introduction**

The natural life cycle of budding yeast alternates between haploid and diploid phases. Both ploidies can be stably propagated asexually through mitotic division. Both theory and experimental work show that haploids adapt faster than diploids, likely due to recessive beneficial mutations (Orr & Otto, 1994; Zeyl, Vanderford, & Carter, 2003). Curiously, however, repeated attempts at evolving experimental haploid populations have resulted in recurrent whole genome duplications yielding populations of autodiploids (Gerstein et al., 2006; Hong & Gresham, 2014; Voordeckers et al., 2015) see **Table 2-1**). Proposed explanations of this phenomenon include artifacts of strain construction (Venkataram et al., 2016), unintended mating events (Voordeckers et al., 2015), and an adaptive advantage of diploidy (Gerstein et al., 2006).

Whole genome duplication (WGD) in asexual haploid populations could provide a fitness advantage in several different ways. Cell size scales with DNA content across many taxa including yeast (Beaulieu, Leitch, Patel, Pendharkar, & Knight, 2008; Epstein, 1967; Gregory, 2001), and increased cell size may facilitate more rapid metabolism and increased growth rate. Indeed, increased cell volume has been reported in laboratory-evolved microbial populations (Lenski, & Travisano, 1994). Gene expression patterns also vary with ploidy (Galitski, Saldanha, Styles, Lander, & Fink, 1999), and diploid-specific gene regulation may be optimal. “Ploidy drive” has been used to describe the phenomenon by which ploidy changes in evolving fungi favor restoration of the historical ploidy state (Gerstein, Lim, Berman, & Hickman, 2017).

**Table 2-1.** Observations of autodiploidy in experimental studies

<b>Study</b>	<b>Propagation</b>	<b>Evolution medium</b>	<b>Strain background</b>	<b>Mating-type</b>
Current study	Batch culture, unshaken	YPD	W303	<i>MATa</i> & <i>MATα</i>
Kosheleva and Desai 2017	Batch culture, unshaken	YPD	Sk1-W303 hybrid	<i>MATa</i> & <i>MATα</i>
Gorter 2017	Batch culture, shaken	YPD with heavy metals	BY4743	<i>MATa</i>
Venkataram <i>et al.</i> 2016	Batch culture, shaken	Carbon limited glucose	BY4709	<i>MATa</i>
Voordeckers <i>et al.</i> 2015	Turbidostat	6-12% EtOH glucose	S288c derivative	<i>MATα</i>
Hong and Gresham 2014	Chemostat	Nitrogen limited glucose	S288c derivative	<i>MATa</i>
Oud <i>et al.</i> 2013	Anaerobic batch culture in sequential bioreactor	1:1 glucose/galactose	CEN.PK113-7D	<i>MATa</i>
Gerstein <i>et al.</i> 2006	Batch culture, shaken	YPD	SM2185	<i>MATa</i>

Natural *Saccharomyces cerevisiae* isolates are typically diploid (Liti, 2015) and occasionally polyploid (Ezov et al., 2006). If most selection has occurred on these higher ploidy states, then gene regulation and cell physiology of diploids should be better optimized relative to haploids.

Despite the recurrence of diploidization events in haploid-founded yeast lineages, the nature of the fitness advantage of diploidy remains unclear. Some studies detect a fitness benefit (Gorter et al., 2017; Venkataram et al., 2016), while no advantage is detected in others (Gerstein & Otto, 2011; Hong & Gresham, 2014). A survey of the effect of ploidy on growth rate in otherwise isogenic strains indicates that the benefit of ploidy varies across conditions and optimal ploidy states are contingent on environment (Zörgö et al., 2013). In environments where duplication does not confer a direct fitness advantage, it may afford indirect benefits that are then themselves acted upon by selection. Diploidy may transiently protect evolving lineages from purifying selection by masking the effects of deleterious recessive mutations over short time scales. Indeed, 15% of viable single gene deletions in haploids exhibit growth defects in rich media, while 97% of heterozygous gene deletions show no detectable phenotype in the absence of perturbation (Deutschbauer et al., 2005). This “masking” hypothesis also has experimental support from mutagenesis studies (Mable & Otto, 2001), and this effect could be advantageous in populations in which the deleterious mutation rate is sufficiently high.

Autodiploids could invade haploid populations due to increased access to beneficial mutations. Ploidy-dependent mutations are known to arise in experimental evolution (Gerstein et al., 2013; Marad et al., 2018), and a favorable shift in the



distribution of fitness effects may follow genome duplication. Structural variants - deletions, amplifications, and translocations - have repeatedly been shown to be adaptive in experimentally evolving yeast populations (Dunham et al., 2002; Gresham, David et al., 2008). Diploids have a greater tendency to form copy number variants (CNVs), especially large deletions (Zhang et al., 2013). Likewise, aneuploidies accumulate at a significantly higher rate in diploids in the absence of selection (Sharp, Sandell, James, & Otto, 2018). If structural variants are more frequent, more variable, and more tolerable in diploids, genome duplication may enable access to novel adaptive paths. Given the repeated observation of displacement of haploids by diploids (**Table 2-1**), and the absence of clear evidence for instantaneous fitness advantages of isogenic diploidy that is broadly applicable across experiments, it is possible that selection for and maintenance of diploidy is a complex process involving both direct selection on ploidy state and second order selection, or selection for indirect fitness benefits associated with higher ploidy.

Here we show recurrent WGD in 46 haploid-founded populations during 4,000 generations of laboratory evolution in rich media. We track the dynamics of genome duplication across the haploid-founded populations, revealing that autodiploids fix by generation 1,000 in all 46 populations. Competitive fitness assays show that WGD provides a 3.6% fitness benefit in the selective environment. We find that the immediate fitness gain is accompanied by a loss of access to recessive beneficial mutations. As a consequence, the rate of adaptation of autodiploids slows. Sequencing of the evolved genomes indicates that autodiploids have increased access to structural variants and largely utilize a different spectrum of mutations to adapt compared to haploids. Finally, we show that autodiploids are buffered from the effects of recessive deleterious

mutations, consistent with an initial benefit to a newly formed diploid genome and loss of redundancy following WGD.

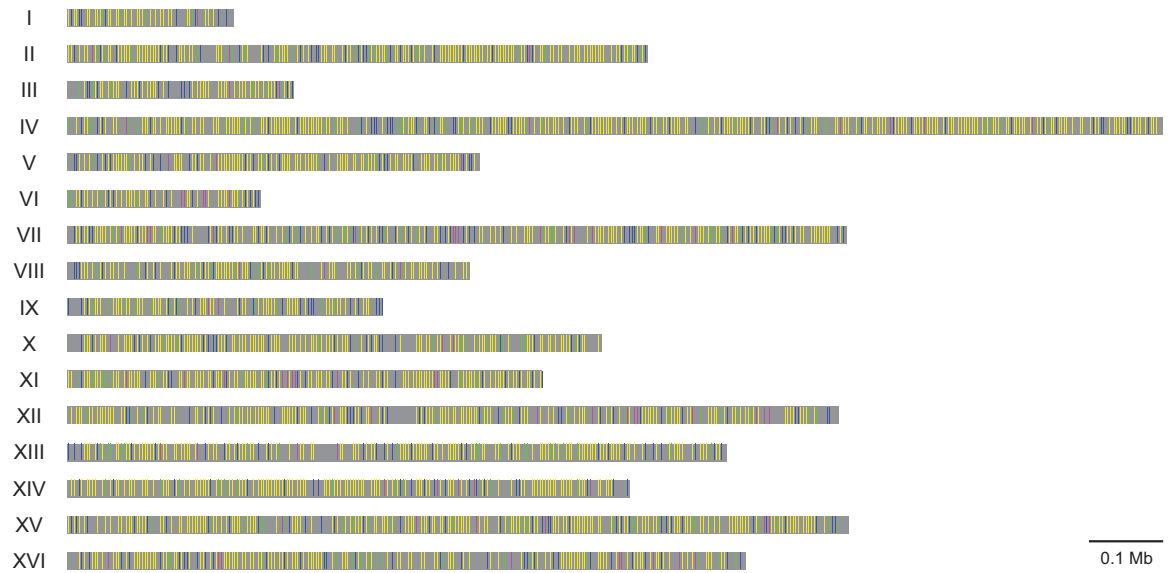
## **Results**

### **Sequenced genomes indicate early and recurrent fixation of autodiploids**

Two clones were sequenced from each of 46 haploid-founded populations after 4,000 generations of evolution, revealing over 5,100 *de novo* mutations distributed uniformly across the genome, representing the largest dataset of mutations identified in *S. cerevisiae* experimental evolution to date (**Fig. 2-1; Dataset 1**). Mutations are normally distributed across clones (one-sample Kolmogorov-Smirnov test,  $\alpha=0.05$ ) with a mean of  $91 \pm 20$ . Most mutations in the sequenced clones were called at  $\sim 0.5$  (implying heterozygosity), a surprising result given that a haploid ancestor founded the populations. Recurrent WGD events were suspected given that each clone maintained its ancestral mating-type allele. Further, this hypothesis of WGD was supported by the observation that clones are not heterozygous at the 6 polymorphic sites that differ between the *MATa* and *MAT $\alpha$*  strains. Finally, evolved autodiploids are mating competent, pointing to duplication of haploid genotypes.

### **Autodiploids are detected early, sweep quickly, and exhibit a fitness advantage**

We determined the fitness effect of genome duplication by directly competing *MATa/a* autodiploids against an otherwise isogenic haploid *MATa* reference. To control for possible artifacts of construction, we independently constructed and competed 10 *MATa/a* diploids. All 10 *MATa/a* autodiploid reconstructions exhibit a relative fitness advantage significantly higher than a control haploid strain (Welch's t-test,  $t=16.28$  *df*

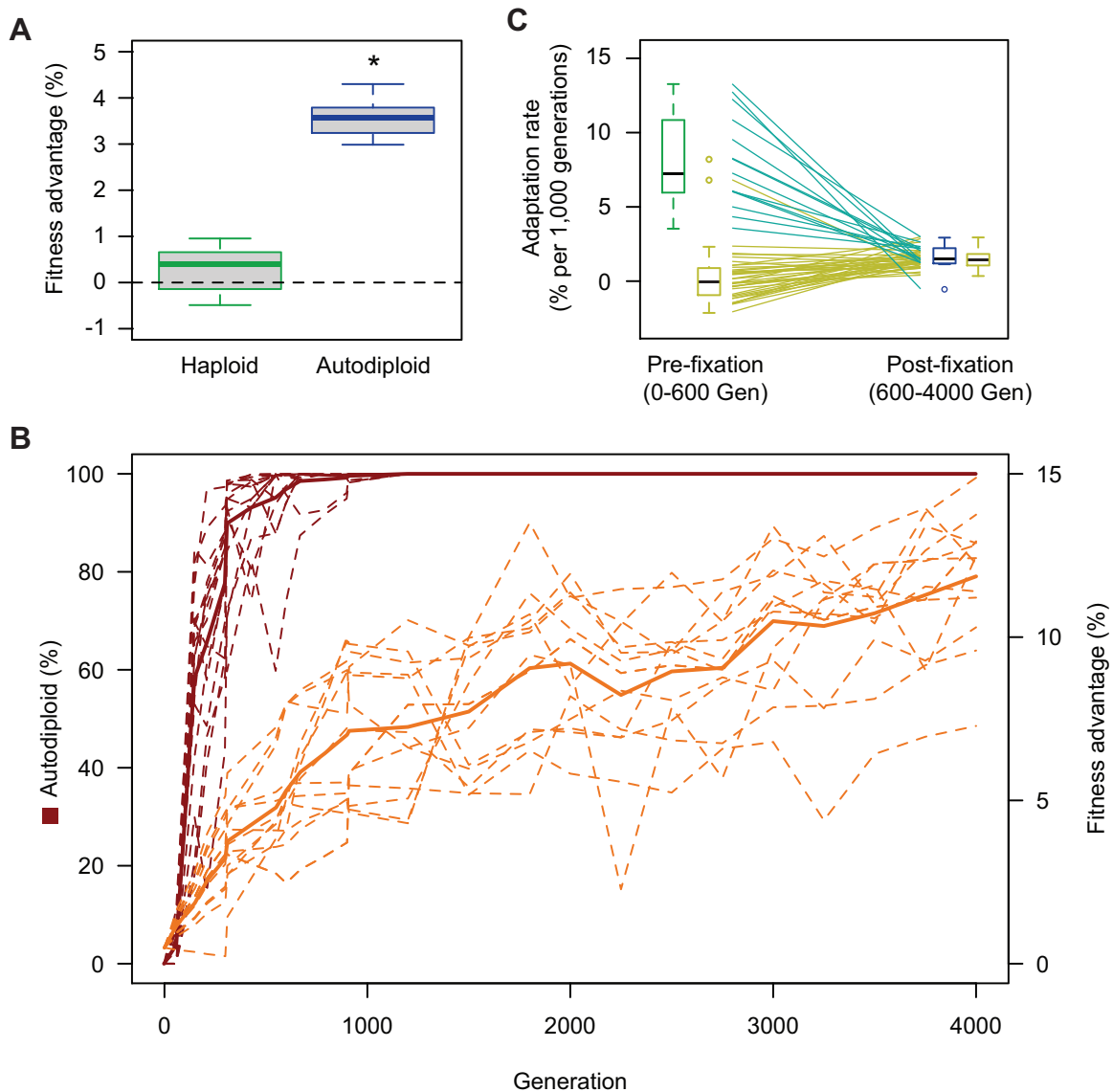


**Figure 2-1.** Colored bars denote the genomic positions of all 5,016 evolved mutations identified in this experiment across the 16 yeast nuclear chromosomes. Evolved mutations are distributed evenly across the genome. Mutations are colored by type: nonsynonymous – yellow, synonymous – green, intergenic – blue, tRNA – magenta.

=19,  $p < 0.001$ ). Genome duplication alone in the absence of any other variation provides a mean fitness benefit of 3.6% in these experimental conditions (**Fig. 2-2A**).

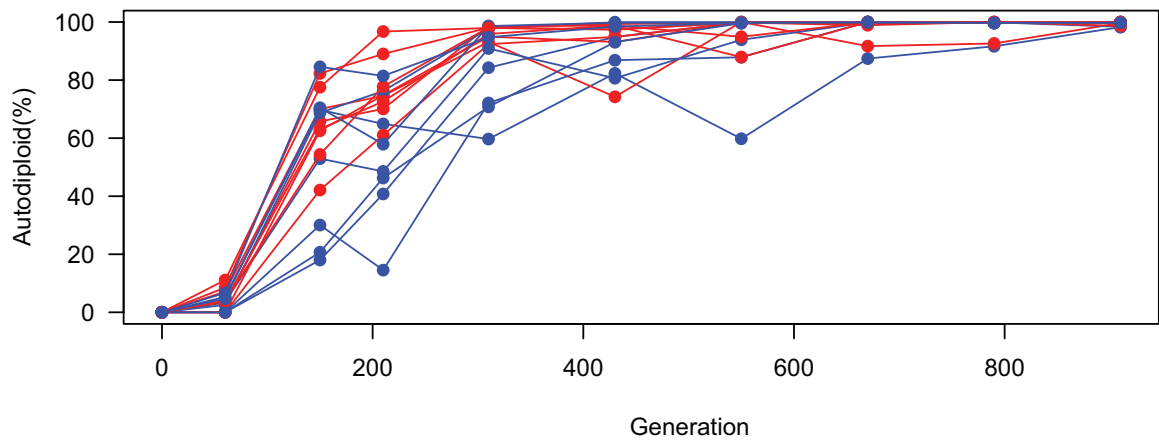
To determine the timing of duplication events, we performed time-course DNA content staining on cryoarchived samples for 16 randomly selected populations (8 of each mating-type). Autodiploids arise quickly in all 16 populations, fixing by generation 1,000 in all but 2 populations (**Fig. 2-2B, Fig. 2-3, Fig. 2-4**). Diploids are present at 2% - 11% in 11/16 populations at generation 60, the earliest time point available for assay. Some populations appear to show clonal interference by fit haploids, with autodiploid fractions briefly decreasing between some time points. Aside from such slight variations, patterns of emergence and spread of autodiploids display similar dynamics for all 16 populations examined.

We examined whether the degree of parallelism observed in ploidy dynamics can be attributed to ancestral ploidy polymorphisms present at the onset of the experiment. Four lines of evidence support the independent origin of autodiploidy in this experiment. First, the cultures were initiated from two starting strains (*MATa* and *MATα*). There is no significant difference in autodiploid frequency between mating-types at any generation (**Fig. 2-3**), meaning if autodiploids did, in fact, arise in both independent inoculating cultures, they would have had to achieve roughly the same frequency, which is highly unlikely. Second, no diploids were detected by DNA content staining in any populations at Generation 0, indicating autodiploids were not present in the inocula above our detection limit of 1%. Third, computational simulations show that low frequency autodiploids are insufficient to explain the recurrent observation of autodiploid fixation events in all 46 replicate populations. Autodiploids with a 3.6% fitness advantage starting

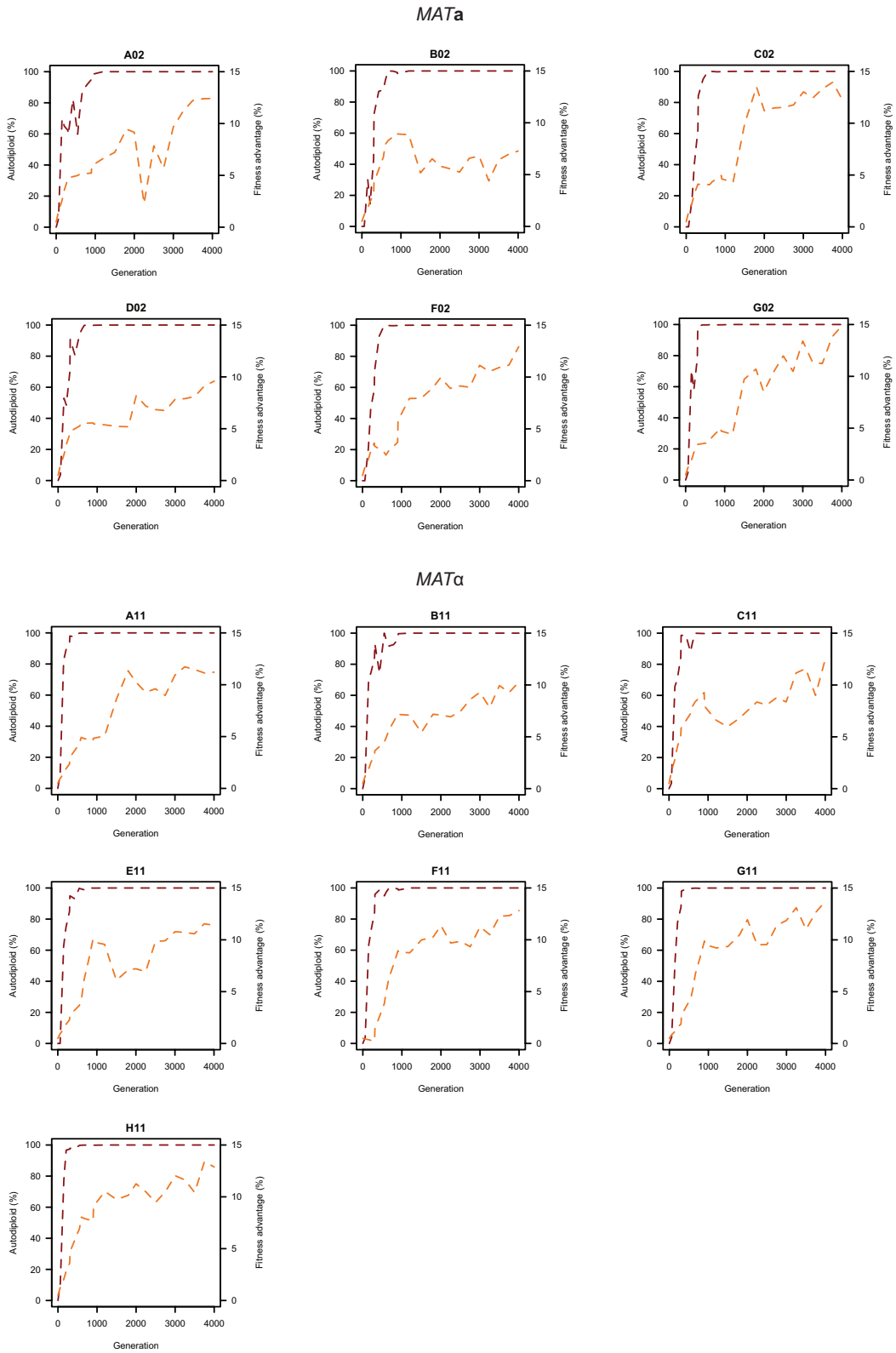


**Figure 2-2.** **A)** *MATa/a* diploids have a mean relative fitness advantage of 3.6% when competed against a haploid reference strain. Ten *MATa/a* diploids clones were constructed independently. Box plots reflect mean fitness of each clone. Autodiploids and control haploids were competed against the same haploid reference. \*  $p < 0.001$  (Welch's t-test) **B)** Autodiploid frequency (red) and fitness advantage (orange) for focal populations (dashed lines). Solid lines indicate mean autodiploid frequency for 16 populations and mean fitness advantage for 13 populations. **C)** Haploid-founded populations demonstrate significantly higher rates of adaptation until autodiploids fix. From that point forward, haploid-founded (autodiploids) and diploid-founded populations adapt at the same rate. Lines indicate paired data points from the same population (teal: haploid-founded, yellow: diploid-founded). For each haploid-founded population, adaptation rate was calculated before and after autodiploid fixation, which occurred on

average at generation 600. Adaptation rates for diploid-founded populations (diploid data reported in Marad et al. 2018) were calculated from Gen 0–600 and Gen 600–4000.



**Figure 2-3.** Autodioids were tracked in 16 focal populations via time-course DNA content staining. Autodioid lineages arise quickly in all 16 populations and fix by generation 1,000 in all but 2 populations. *MATa* (n = 8) and *MATα* (n = 8) are represented by red and blue lines, respectively.



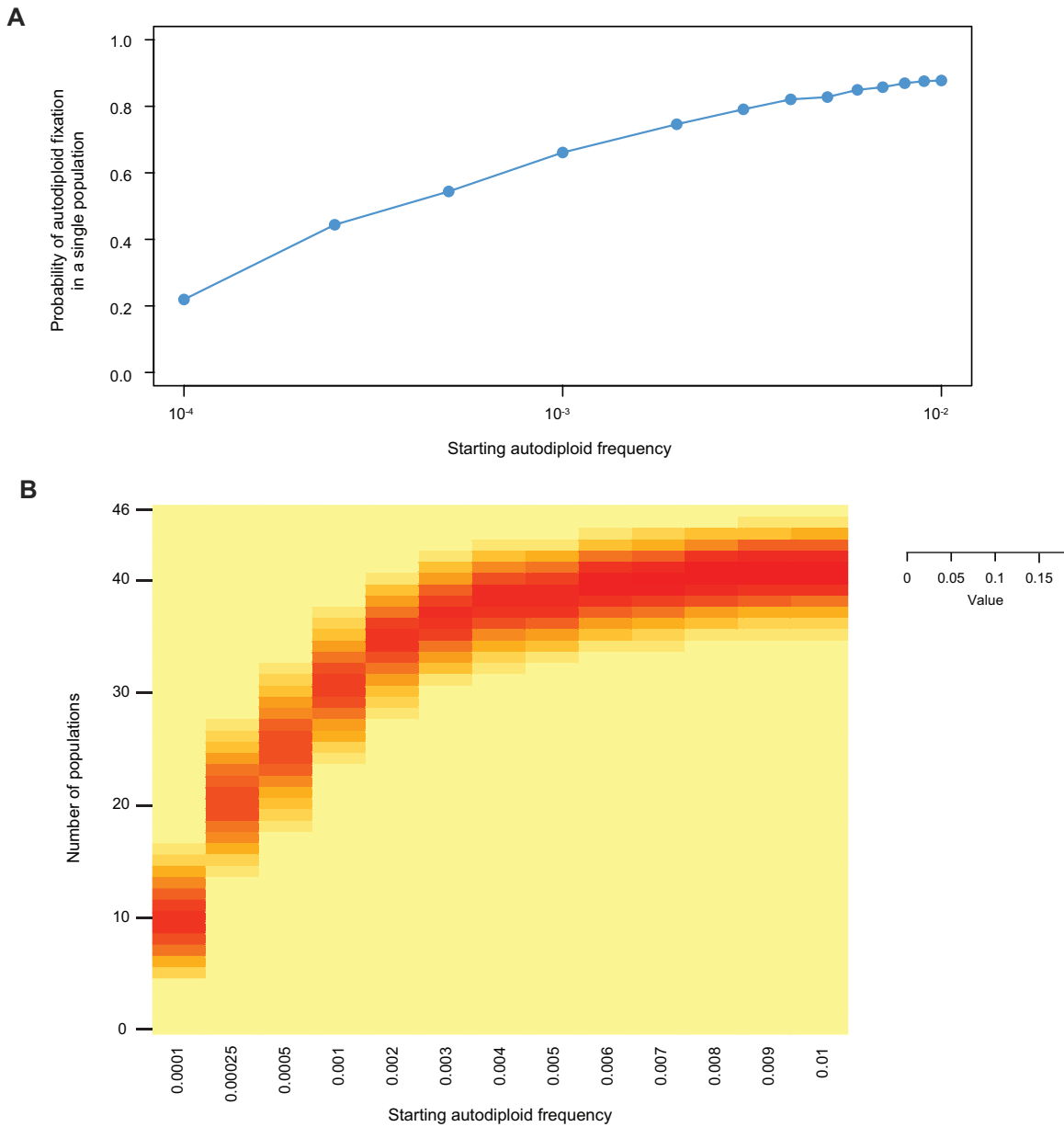
**Figure 2-4.** Time-course ploidy (red) and fitness (orange) dynamics across 4,000 generations for the 13 populations for which both have been measured.



at a frequency of 0.01, the highest frequency we modeled, have a probability of fixation in a given population of 0.88 and therefore the chance of fixation in all 46 populations would be  $2.5 \times 10^{-3}$  (**Fig. 2-5A-B**). A fourth line of evidence is the recent reporting of a high rate of autodiploid occurrence in passaged yeast cultures. Harari *et al.* (2018) report a rate ploidy transition on the order of  $10^{-5}$  per cell division, which corresponds to hundreds of WGD events generated during each 24-hour growth cycle. Taken together, this argues that, while ancestral autodiploids may have swept in some populations, ancestral ploidy variation is insufficient to explain autodiploid fixation in all 46 populations. Therefore independent, parallel WGD events during the evolution experiment are necessary to explain the recurrent fixation reported here.

### **Autodiploids adapt more slowly than haploids**

Consistent with previous work (Gerstein, Cleathero, Mandegar, & Otto, 2011; Marad et al., 2018), we find that WGD in haploids provides an immediate fitness gain at the expense of slowing subsequent adaptation. To examine how the shift to diploidy impacted the dynamics of adaptive evolution, we measured population fitness for all populations at ~300-generation intervals. Mean time-course fitness estimates show a change in slope following 1,000 generations. This corresponds roughly to the time that autodiploids have fixed in most focal populations and are high frequency in the remaining populations (**Fig. 2-2B**). We compared the rate of adaptation before and after the fixation of diploids in 13 focal populations for which quality fitness data was available. Because many factors, including epistasis, could explain a change in adaptation rate over time, we used a repeated measures ANOVA to compare the effect of ploidy on adaptation rate using time-course fitness data from diploid-founded populations that were

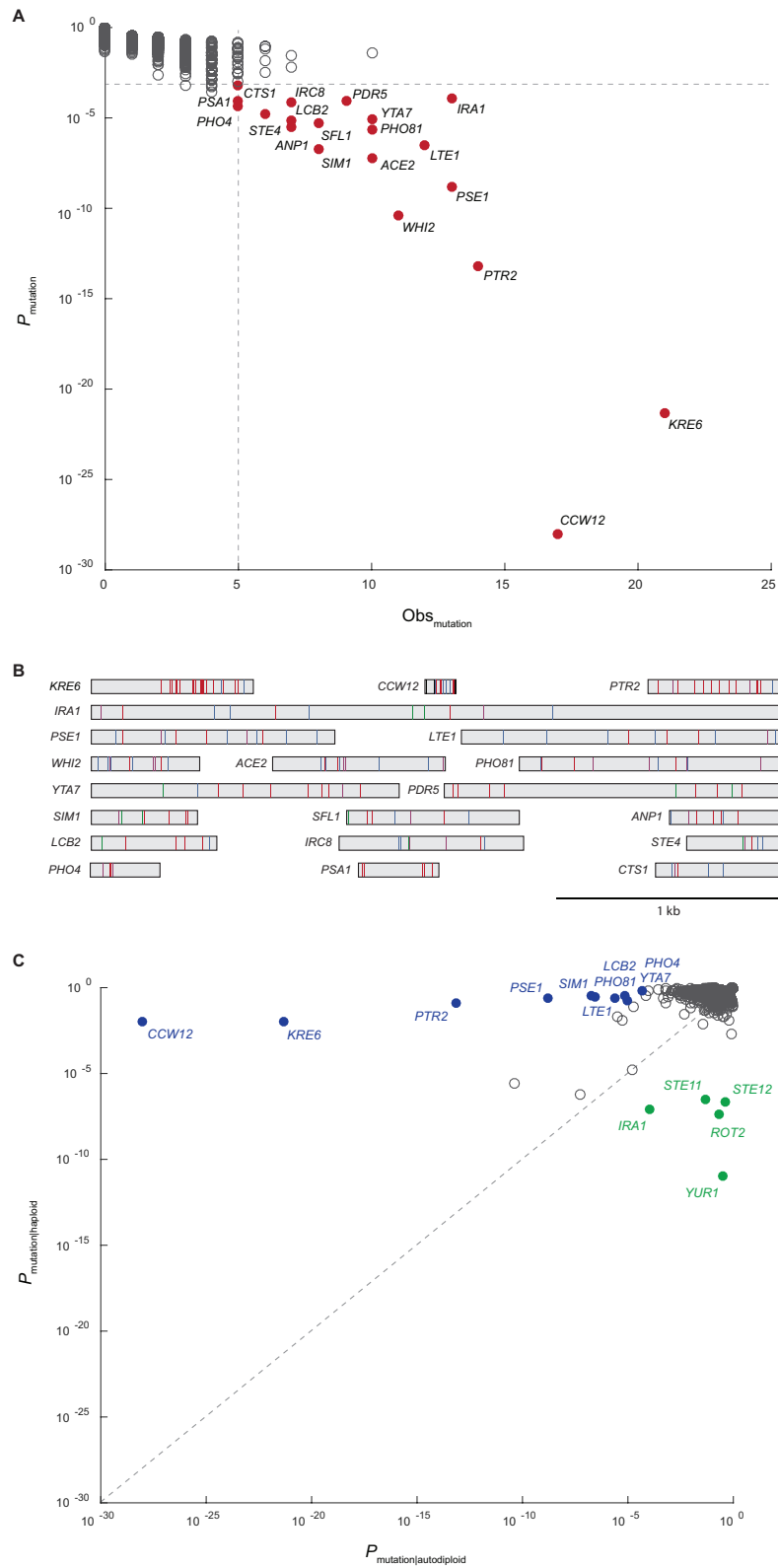


**Figure 2-5.** **A)** The simulation derived probability of autopolyploid fixation at starting frequencies ranging from 0.0001 to 0.01. Each data point represents the proportion of populations that fix autopolyploids in 10,000 simulations. **B)** Heatmap showing the probability distributions of autopolyploid fixation at a given starting frequency.

evolved in parallel (Marad et al., 2018) (**Fig. 2-2C**). The interaction of founding ploidy and generation has a significant effect ( $F(1, 49)=78.04, p<0.001, \eta_p^2 = 0.614$ ). Post hoc comparisons using a Bonferroni correction indicate that rates of adaptation are significantly higher in haploid-founded populations than diploids ( $p<0.001$ ), and that adaptation rate does not differ once autodiploids fix ( $p=0.38$ ). Duplication itself is a significant component of incipient haploid adaptation, however, diploidy alone is unable to account for the range of population fitness values at the time point in which diploids fix, which ranges from 1.9% to 8.0%. Therefore, additional beneficial mutations are needed to explain high gains in fitness in some populations.

### **Autodiploid genomes harbor autodiploid specific mutations**

Duplication of a haploid genome affects both cell physiology and the phenotypic consequences of new mutations. Therefore, the selective pressure on a gene may vary depending on ploidy state. To understand how genome evolution is driving adaptation in the autodiploid populations, we utilize a recurrence approach that accounts for both the number of mutations observed in a gene and the expectation that the observed number of mutations of a given gene occurred by chance alone controlling for gene length. The resulting probabilities were used to identify 20 common genic targets of selection (**Fig. 2-6A, Table 2-2**). There is a median of four recurrent targets per clone with only one population containing no common target mutations. GO-component term analysis indicates common targets are enriched for genes whose protein products localize to the cell periphery ( $p = 0.001$ ). Cell periphery targets include *CCW12* and *KRE6*, which both appear to be under extremely strong selective pressure when using the probability metric as a proxy for strength of selection. Interestingly, a tRNA gene, tL(GAG)G, was also



**Figure 2-6. A)** The observed number of coding sequence mutations in each of the 5800 genes in the S288c reference genome plotted against the probability that the observed

number of mutations in each gene occurred by chance. Common targets of selection (solid red circles) are genes with 5 or more coding sequence mutations and a corresponding probability of less than 0.1%. **B)** Shown are all 188 mutations across the 20 common targets of selection. Mutations are colored by type: frameshift-purple, nonsense-blue, missense-red, synonymous-green, other-black. Both homozygous and heterozygous mutations are shown. **C)** Plotted is the probability that the observed number of CDS mutations in a gene occurred by chance in haploid populations (haploid data reported in (Lang et al., 2013) versus autodiploid populations. Haploid-enriched genes are indicated by solid green circles and autodiploid-enriched genes as solid blue circles.

**Table 2-2. Mutations occurring in common target genes.**

Gene	Number of Mutations (Populations)			Evolved Alleles	GO Biological Process Term <sup>3</sup>
	Hom	Het	Mixed <sup>2</sup>		
<i>KRE6</i>	1	19	1	T311R, A351T, D361H, T376N, R380G, S393L, W447L, W447L, C454F, S464Y, P487S, G492D, E497D, D499A, S517F, N545K, Y579*, G590D, W642C, D654G, Q681*	Fungal-type cell wall organization
<i>CCW12</i>	1	14	0	M1Startloss, S39del, C40F, E41*, S50C, D62Y, Q67*, Q67*, Y68fs, T70N, E77*, E93*, Y110*, L125S, G127S	Fungal-type cell wall organization
<i>PTR2</i>	2	12	0	A43A, G110fs, G128R, M203T, V243F, C279F, W313S, P359R, A391P, Y452fs, S484Y, A491G, K500E, Y555*	Peptide transport
<i>IRA1</i>	2 <sup>1</sup>	11	0	L37F T39fs N137K, Q550*, S622*, T820M, L974*, I1437I, F1489F, S1603G, S1753I, C1754fs, C2067*	Negative regulation of Ras protein signal transduction
<i>PSE1</i>	0	12	0	L107*, W137S, Q308fs, W331*, L372V, I517L W606*, V697fs, Q739*, E765Q, L869*, S1006*	Protein import into nucleus
<i>WHI2</i>	6 <sup>1</sup>	3	2 <sup>1</sup>	Q29*, S72*, L76fs, L76fs, Q81*, E168G, Q181*, N275fs, T283fs, A310P, A338*	Regulation of growth
<i>LTE1</i>	0	11	0	S185*, W380*, E653*, A748S, A748V, E865*, Q916K, M1062I, K1138*, A1368fs, W1403*	Regulation of exit from mitosis
<i>YTA7</i>	0	9	1	S319S, E475*, P564H, P675Q, L803F, L965F, I1032S, P1061R, R1120fs, A1203P	Negative regulation of transcription
<i>PHO81</i>	4 <sup>1</sup>	6	0	R93*, F96L, N244N, N329fs, A582fs, E621*, P699Q, L748fs, PR753*, V1079fs	Phosphate-containing compound metabolic process
<i>ACE2</i>	2	7	1	E213*, R227H, E235*, P288Q, S299*, P314fs, N324fs, S473*, S694*, L770fs	Positive regulation of cell separation after cytokinesis
<i>PDR5</i>	2	7	0	T39I, Q56L, S197F, A262S, D1035D, N1120K, F1224Y, V1290V, S1331Y	Drug transport
<i>SFL1</i>	3 <sup>1</sup>	5	0	E4E, G88R S114R, S213*, S283fs, P432H, DY544*, Y545*	Negative regulation of invasive growth
<i>SIM1</i>	0	8	0	A119fs, L132L, G222G, G234C, V235G, P344L, L418W, A427T	Fungal-type cell wall organization
<i>IRC8</i>	2 <sup>1</sup>	5	0	L262*, Q274*, L310L, N316fs, D474fs, Q629E, L649*	Mitotic recombination
<i>LCB2</i>	0	7	0	H44H, F148fs, T149I, G373C S414N, R494T, S526*	Sphingolipid biosynthetic process
<i>ANP1</i>	0	7	0	K2*, R82fs, S120C, P195Q, V230L, IQ241*, G303W	Protein N-linked glycosylation
<i>CTS1</i>	5 <sup>1</sup>	0	0	L69*, L83fs, C96Y, Q234*, E298*	Cell separation after cytokinesis
<i>PSA1</i>	0	5	0	R15K, P24A, G284S, L293S, D330T	Cell wall mannoprotein biosynthetic process
<i>PHO4</i>	0	5	0	F171F, D175fs, L270V, V286M, A298fs	Positive regulation of phosphate metabolic process
<i>STE4</i>	0	5	0	G250G, S261fs, A287S, R312Q, E315*, Q337*	Pheromone-dependent signal transduction

<sup>1</sup>Includes mutations putatively homozygous with inconclusive coverage. <sup>2</sup>Mixed mutations are present as both homozygotes and heterozygotes in the population. <sup>3</sup>GO terms were manually curated using the Yeast Genome Database (yeastgenome.org).

identified as a common target of selection. This is the first evidence of adaptive tRNA mutations in laboratory yeast evolution. To better understand the molecular basis of adaptation, we examined the distribution of mutations within each gene (**Fig. 2-6B**). Three broad patterns emerge. First, we observe selection for loss-of-function alleles, e.g. 9 of 11 mutations in *WHI2* are high impact (frameshift or nonsense). Adaptive loss-of-function alleles are common in experimental microbial evolution (Cooper et al., 2001; Kvitek & Sherlock, 2013; Venkataram et al., 2016). We also observe selection for change-of-function alleles. For example, only missense and synonymous mutations are seen in *PDR5*. Finally, we observe mutations in common targets that cluster within specific domains. This is illustrated by the clustering of mutations in the C-terminus of both *KRE6* (n=21) and *STE4* (n=6).

We compared the common targets of selection identified in autodiploid clones to those identified with the same approach in a comparable haploid dataset (Lang et al., 2013). We identify several haploid- and autodiploid-enriched targets (**Fig. 2-6C**). Ploidy-enriched targets include genes mutated more often in one ploidy (e.g. *CCW12* and *KRE6* in autodiploids; *YURI* and *ROT2* in haploids) or exclusively in one ploidy (e.g. *PHO81*, *YTA7*, *IRC8* in autodiploids; *STE12* in haploids).

### **Loss of heterozygosity hotspots occur on Chromosomes XII and XV**

Though most mutations are heterozygous, clones contain up to 17 homozygous mutations, with an average of 5.4. Homozygous mutations could either represent mutations that arose before duplication events or loss of heterozygosity (LOH) of heterozygous mutations. We find that the homozygous mutations are not distributed randomly throughout the genome; instead, they tend to cluster in particular regions of the

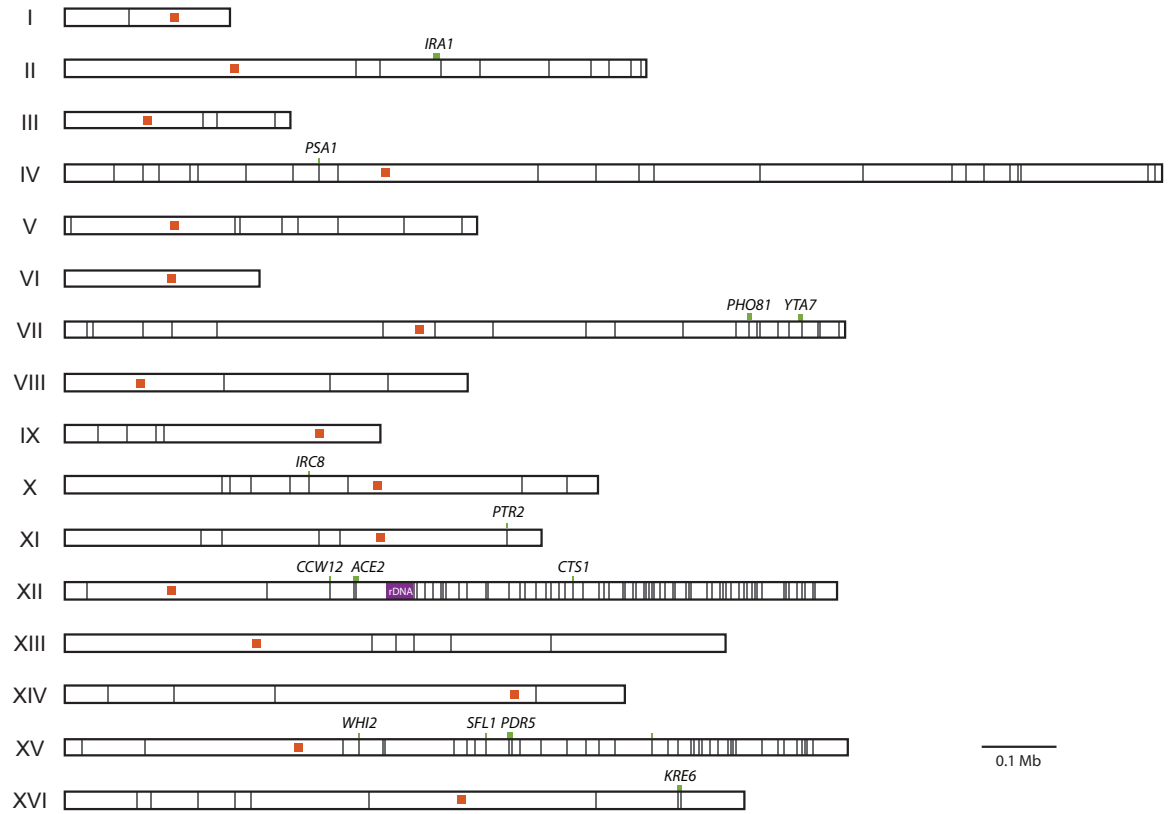
genome (**Fig. 2-7**). These clusters, located on the right arms of Chr. XII and Chr. XV, account for 55% of all homozygous mutations. This clustering implies that most homozygous variants result from recombination events. By removing homozygous mutations occurring in these regions from analysis, the average number of homozygous mutations per clone drops to 2.4. This confirms that only a few mutations arose in a haploid background and that most genome evolution occurred post genome duplication.

Mutations in the common targets of selection are observed as both homozygous and heterozygous. Most genes (12/20) are found mutated in both heterozygous and homozygous states across clones, indicating partial or full dominance of fitness effects. Seven genes only ever contain heterozygous mutations (*ANPI*, *LCB2*, *LTE1*, *PHO4*, *SIMI*, *STE4*, *PSE1*). These mutations are candidates for overdominant effects (Sellis, Callahan, Petrov, & Messer, 2011). Finally, only one gene, *CTS1*, is never found mutated in a heterozygous state. A reasonable hypothesis would be that the *cts1* mutations are recessive; however, we have previously identified *cts1* mutations in evolved diploid populations and found it to be close to fully dominant (Marad et al., 2018). Instead, the position of *CTS1* on the right arm of Chr. XII, a LOH hotspot, could explain why it is only observed in a homozygous state (**Fig. 2-7**).

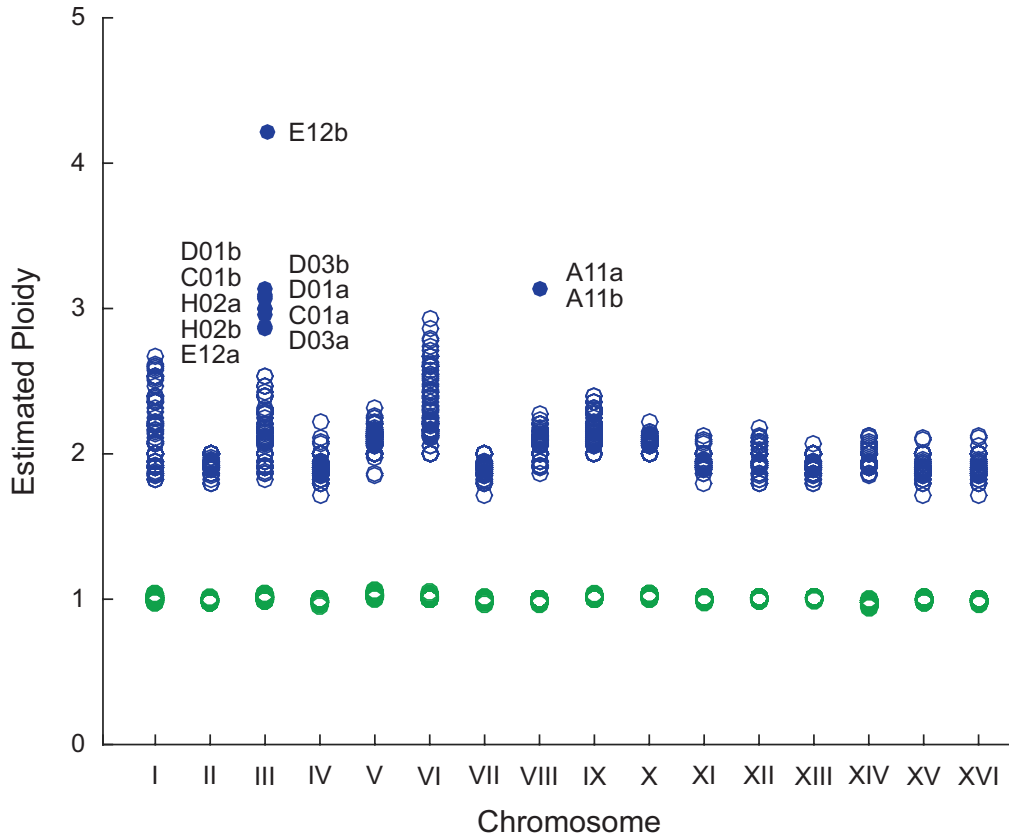
### **Structural variants are common to autodiploids**

In addition to changing the genetic targets of selection, genome duplication permits access to structural variants not accessible to haploid genomes. We analyzed aneuploidies and copy number variants (CNVs) in autodiploid genomes as well as previously sequenced haploid populations (Lang et al., 2013) (**Figs. 2-8; Datasets 2 & 3**). Two types of aneuploidies are observed in autodiploids: trisomy III (which fixes in





**Figure 2-7.** The loci of all homozygous evolved mutations across the 16 yeast nuclear chromosomes are indicated with solid lines. Mutations in common targets of selection are labeled with gene names. Red squares indicate centromere location. Homozygous mutations are not distributed evenly across the genome and cluster on the right arms of chromosomes XII and XV.



**Figure 2-8.** Coverage across each chromosome was compared to genome-wide coverage for each sequenced clone. Euploidy is indicated by empty circles: haploid—green, autopolyploids—blue. Aneuploidies are shown as filled circles and labeled by clone.

five populations) and trisomy VIII (which fixes in one) (**Table 2-3**). CNVs are common in autodiploid genomes. Of the 46 autodiploid populations, CNVs appear in 19 and fix in 14. The 19 independently occurring autodiploid CNVs fall into 10 groups based on genomic position. Autodiploid CNVs consist of both amplifications (n=4) and deletions (n=6). In contrast, no aneuploidies and only two amplifications are detected amongst the 40 haploid populations. These two amplifications are also observed in autodiploids.

### **Autodiploids are buffered from deleterious mutations**

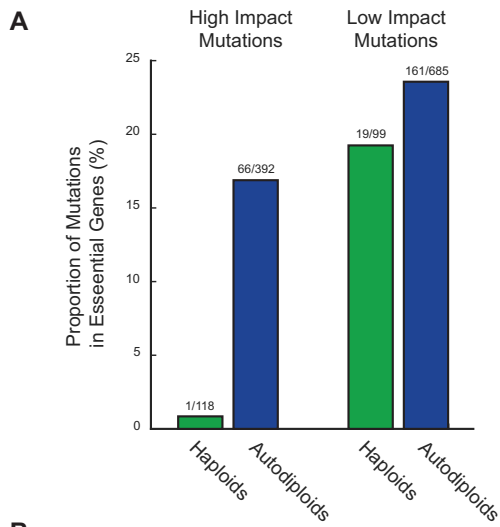
To determine the extent to which an increase in ploidy buffers diploid lineages against the effects of deleterious mutations, we compared the frequency of mutations in essential genes in autodiploids with those of *MATa* haploids described previously (Lang et al., 2013). We specifically analyzed frameshift and nonsense mutations that would likely phenocopy the null mutants used to characterize genes as essential. Sixty-three of 66 high impact mutations in essential genes are heterozygous. For the remaining three mutations, zygosity is inconclusive due to low coverage. We find high impact mutations in essential genes to be exceptionally rare in haploids, with only a single case observed (**Fig. 2-9A**). In contrast, autodiploids contain a significantly higher proportion of high impact mutations in essential genes ( $\chi^2(1) = 20.32, p < 0.0001$ ). As expected, the proportion of low impact mutations within essential genes is consistent across ploidies ( $\chi^2(1) = 0.909, p = 0.339$ ). Essential genes are also present within two of the large deletions observed in autodiploids (**Table 2-3**).

To experimentally validate that recessive lethal mutations accumulate in autodiploids, we sporulated three *MATa/a* from three different populations and performed tetrad dissections. Clones A02a, B01a, and C03b were selected because they contain no

**Table 2-3. Copy number polymorphisms in evolved clones.**

Chr	Start (kb)	End (kb)	Length (kb)	Copy Number	Class	Type	Clones*
I	210	225	15	1N	CNV	loss	<b>B01a, B01b, E11a, E11b</b>
III	85	85	<10	0N	CNV	loss	<b>G01a, G01b, G01c</b>
III	150	170	20	1N	CNV	loss	<b>A02a, A02b, B10a, B10b, C11a, C11b, C11c, F10a</b>
IV <sup>2</sup>	900	1000	100	3N	CNV	gain	<b>B12a, B12b, C03a, E12a, E12b</b>
V <sup>3</sup>	450	500	50	1N	CNV	loss	<b>B11a, B11b, F10a, F10b</b>
VIII	525	545	20	1N	CNV	loss	<b>E11a, E11b</b>
XIII <sup>3</sup>	190	200	10	1N	CNV	loss	C10a, D10a, E10c, H12a
XIII <sup>2</sup>	190	200	10	3N	CNV	gain	<b>F02a, F02b</b>
XIV	545	560	15	3N	CNV	gain	<b>A12a, A12b</b>
XV	900	1100	200	3N	CNV	gain	G02b
III	0	317	317	3N <sup>1</sup>	aneuploidy	gain	<b>C01a, C01b, D01a, D01b, D03a, D03b, E12a, E12b<sup>1</sup>, H02a, H02b,</b>
VIII	0	924	924	3N	aneuploidy	gain	<b>A11a, A11b</b>

\*Bolded clones indicate the CNV was found in all clones of the population. <sup>1</sup>Observed at 4N in one clone. <sup>2</sup> Also observed in one haploid. <sup>3</sup> Contains essential genes.



**B**

Clone	High impact CDS mutations in essential genes	Predicted spore viability	Measured spore viability	4 spore viable	3 spore viable	2 spore viable	1 spore viable	0 spore viable	Sporulation efficiency (72h)	Small colony size
A02a	1	50%	3.75%**	0	0	0	3	17	36.3%	2
B01a	0	100%	66.25%*	5	7	5	2	1	11.9%	18
C03b	2	25%	17.5%	0	0	1	12	7	46.5%	6
Control	NA	100%	96.25%	18	1	1	0	0	44%	0

\*\*  $\chi^2(1) = 43.538, p < 0.00001$   
 \*  $\chi^2(1) = 30.258, p < 0.00001$

**Figure 2-9. A)** Proportions of high impact mutations (frameshift, nonsense) and low impact mutations (synonymous, intronic) in essential genes in haploids (green) and autodiploids (blue). Above each bar is the ratio of mutations in essential genes to mutations in all genes. **B)** Clones from three evolved diploid populations were sporulated and dissected. Spore viability and small colony size reflect recessive lethal and recessive deleterious mutations, respectively.

identifiable aneuploidies that would complicate measures of spore viability. Out of 20 total dissected tetrads (80 total spores) per clone, spore viability ranged from 4% to 66% in evolved autodiploid clones (**Fig. 2-9B**). Further, a substantial fraction of germinated spores developed morphologically small colony sizes relative to controls. We compared observed spore viability to expected viability based on the number of high impact mutations in genes annotated as essential. The only clone for which we observed four-spore viable tetrads, B01a, is also the only clone with no predicted recessive lethal mutations. Nonetheless, both A03a and B01a have significantly lower spore viability than expected. This in part may be due a genetic load imposed by segregating deleterious alleles. Consistent with our sequencing data, these data indicate that diploidy permits the accumulation of recessive lethal and deleterious mutations on a relatively short time scale.

## **Discussion**

Whole genome duplications (WGDs) are significant evolutionary events that have profound impacts on genome evolution. Evidence of ancient whole-genome duplication events is found within lineages ancestral to most extant eukaryotic taxa (Jaillon et al., 2004; Meyer & Van de Peer, 2005; Tang et al., 2008), including at least two WGDs in the vertebrate lineage (Dehal & Boore, 2005), and a WGD approximately 100 mya in the *Saccharomyces* lineage (Kellis, Birren, & Lander, 2004; Wolfe & Shields, 1997). In addition, the existence of numerous contemporary polyploid taxa suggests that genome duplication plays a role in short-term adaptive evolution (Van de Peer, Maere, & Meyer, 2009). Genome duplication and polyploidy are also known to increase virulence and aid in stress adaptation in pathogenic fungi (Gerstein et al., 2015). Here, we show that

experimental evolution of haploid *Saccharomyces cerevisiae* results in rapid and recurrent WGD. Clones with duplicated genomes arise early in all 46 populations and fix rapidly. We show that concurrent fixation of autodiploids can be attributed to a large fitness effect. Furthermore, the concurrent population dynamics reported here are evidence of a high rate of genome doubling in haploid yeast.

The invasion and subsequent fixation of autodiploids in haploid-founded lineages has been reported before in yeast (see **Table 2-1**). Some studies report a fitness advantage of WGD in haploid yeast (Venkataram et al., 2016), though this is not consistent across studies (Gerstein & Otto, 2011). Such inconsistency is possibly because the benefit of diploidy is condition-dependent (Zörgö et al., 2013). By employing a competitive growth assay, we demonstrate a relatively large fitness effect of a duplicated genome in our selective environment. A 3.6% fitness effect is substantial: in a recent study we quantified fitness effects of over 116 mutations from 11 evolved lineages in the same conditions, and only 9 conferred a fitness benefit greater than 3.6% (Buskirk, Peace, & Lang, 2017). The biological basis of this fitness advantage is unclear. However, there are several strong possibilities. Increased cell size, differential gene regulation, and a diploid-specific proteome (De Godoy et al., 2008; Galitski et al., 1999) may all contribute to the adaptive advantage of diploidy. More generally, environmental robustness is often associated with increases in ploidy (Van de Peer et al., 2009)

The recurrent and remarkably parallel manner in which autodiploids arise and fix points to not only a large fitness effect, but a high rate of occurrence. Our previous work has shown that parallel evolution is evident at the level of genetic pathway and even gene (Buskirk et al., 2017; Marad et al., 2018). However, the extent of the convergence

observed here – where all 46 populations evolve to be autodiploids – is unprecedented in our experimental system. While it cannot be dismissed that some autodiploids were present in the founding inoculum, they are below our 1% detection limit. Autodiploids at this low of a frequency in the inoculum is not sufficient to explain the extent of fixation observed. Simulations indicate the probability of an autodiploid lineage at 1% fixing in 46 out of 46 replicate populations is  $2.5 \times 10^{-3}$ . Furthermore, given the common dynamics observed in populations of both mating types, autodiploids would have to have arisen in “jackpot” fashion and reach a similar frequency in the inocula of both mating-types. These data strongly support independent WGD events in replicate populations, suggesting a high background rate of duplication. This is consistent with the observation of frequent WGD in mutation accumulation lines (Lynch et al., 2008) but see conflicting findings using a different strain in (Sharp et al., 2018). Using a barcode-enrichment assay, Venkataram *et al.* (2016) found that roughly half of all evolved clones with increased fitness that arose in a short-term enrichment experiment possessed no mutation apart from a WGD. A recent study found autodiploids to occur in haploid cultures at a rate on the order of  $10^{-5}$  per cell division (Harari, Ram, Rappoport, Hadany, & Kupiec, 2018), a rate several orders of magnitude higher than the per base pair mutation rate and sufficiently high to explain repeated autodiploid appearance in this and other haploid-founded evolution experiments.

Given the prevalence of autodiploids in the present evolution experiment, it is worth asking why autodiploids were not reported in a previous haploid evolution experiment in which ostensibly the identical strain and conditions were used (Lang et al., 2013). It is possible that in the prior experiment autodiploids did not fix or they could



have fixed but were not detected. Despite conscious efforts to maintain identical selective environments, subtle differences in the conditions may exist given that evolution experiments were conducted years apart in different facilities. Indeed, inconsistency in the appearance of WGD across experiments and conditions is common in the field (Gorter et al., 2017; Voordeckers et al., 2015). Even subtle differences in the evolution conditions could shift the selective benefit of autodiploidy and yield population dynamics different from those seen here. Alternatively, it is possible that autodiploids did fix in the previous haploid evolution experiment but went undetected. The populations analyzed in the haploid study were part of a larger ~600 population experiment, and the 40 focal populations were selected based on the presence of a sterile phenotype. Mutations producing sterile phenotypes are predominantly adaptive and recessive loss-of-function (Lang, Murray, & Botstein, 2009). The presence of such beneficial mutations would have biased the selection of populations towards those retaining haploidy. We analyzed a subset of the remaining ~560 populations by DNA content staining and find that ~30% (3 of 10) of them appear autodiploid at generation 1,000, though this is still less frequent than we report here. Further at least one of the forty sequenced populations (RMS1-E09, (Lang et al., 2013) which appeared to be an autodiploid based on the presence of a large number of mutations present at a frequency of 0.5, was confirmed as 2N through ploidy-staining.

The consequences of WGD are apparent on both the phenotypic and genotypic level. One such consequence is the susceptibility of autodiploids to Haldane's sieve, resulting in a "depleted" spectrum of beneficial mutations. We find a decline in adaptation rate following WGD, which mirrors findings from studies that directly

compare the rates of haploid population adaptation with that of diploids (Gerstein et al., 2011; Marad et al., 2018). This implies a fitness tradeoff in the shift from 1N to 2N, wherein the fixation of a large-effect beneficial genotype comes at the cost of eliminating access to future recessive beneficial mutations. This tradeoff associated with genome duplication is predicted when population size is large and most beneficial mutations are partially or fully recessive (Otto, 2007), conditions that are met in our populations (Lang et al., 2013).

Autodiploids share physiological traits with both haploid and diploid cell types. Like their haploid founders, autodiploids possess only a single mating-type allele and will readily mate with cells of the opposite mating-type, indicating haploid-specific regulation of mating-pathway genes. As with diploids, autodiploids possess a 2N genome and exhibit larger cell size (Galitski et al., 1999). Consequently, we observe some overlap in the spectrum of beneficial mutations. We have identified targets of selection shared between haploids and autodiploids along with targets specific to autodiploids. While several targets were mutual to haploids and autodiploids, the extent of recurrence varied by gene. For example, *IRA1* mutations were common to both ploidies but enriched in haploids. In contrast, there were five ploidy-specific genes that were targets in autodiploids but never mutated in haploids. These genes (*PHO81*, *YTA7*, *PHO4*, *IRC8*, and *PSAI*) represent targets of selection that are specifically enriched in autodiploids, suggesting that WGD may expose adaptive pathways that are not easily accessible to either haploids or diploids. The functional basis of selection on a few common genic targets reported here has been investigated (Li et al., 2018; Sezmis, Malerba, Marshall, & McDonald, 2018), and many targets have been observed in evolution experiments before.

However, little is known about the functional consequences of most mutations identified here.

Genome duplication also has consequences on genome stability and the evolution of structural variation. Across our 46 populations we identify 6 independently evolved aneuploidies and 20 independently evolved structural variants. Structural variants are more frequent in autodiploid genomes than in evolved haploid genomes of the same background, even after accounting for length of evolution. Haploids are constrained: whereas the structural variants observed in haploids always result in a net gain of genetic material, autodiploid structural variants include both amplifications and deletions. The ability to generate a greater degree of structural variation could provide a secondary advantage to WGD. Aneuploidies, large rearrangements, and CNVs have been shown to arise and confer an advantage in experimentally evolving yeast populations (Chang, Lai, Tung, & Leu, 2013; Selmecki et al., 2015). Of note, several of the recurrent structural arrangements described in the present study, including trisomy III and a 317 kb deletion on Chr. III, have previously been described as beneficial (Sunshine et al., 2015). The observation of both gain and loss of genetic material from Chr. III may indicate complex selection on phenotypes unachievable through point mutations.

Loss of heterozygosity (LOH) provides a means of overcoming the masking effect of ploidy in autodiploids allowing recessive beneficial mutations to become homozygous. Analysis of the distribution of homozygous mutations across evolved autodiploid genomes reveals LOH frequently occurs in two locations: on the right arm of Chr. XII and the right arm of Chr. XV. The right arm of Chr. XII has been characterized as a hotspot for LOH in experimental and natural populations (Marad et al., 2018;

(Magwene et al., 2011) mediated by a high rate of recombination at the rDNA repeats (Keil & Roeder, 1984). To our knowledge, a mitotic recombination hotspot on Chr. XV has not been described. Recurrent LOH may have substantial evolutionary implications as the affected regions may experience different rates of genome evolution and divergence than the rest of the genome. On the one hand, fitness may decline dramatically due to the exposure of deleterious mutations to selection. On the other hand, the rate of adaptation may be increased by providing access to recessive beneficial mutations that would otherwise be masked by Haldane's sieve. Theory predicts that sufficient mitotic recombination may allow asexual populations to circumvent Haldane's sieve (Mandegar & Otto, 2007). While we only show prevalence of LOH and not functional evidence of adaptive LOH, such events have been repeatedly observed in adapting yeast populations (Gerstein et al., 2014; Smukowski Heil et al., 2017). Further, the LOH on Chr. XV was not detected previously in diploids (Marad., 2018), an observation that is more easily explained by selection than a change in the rate of occurrence.

The same masking effect that stifles recessive beneficial mutations is also predicted to permit the accumulation of deleterious mutations in diploids (Mable & Otto, 2001). In evolved haploid populations few if any deleterious mutations fix: previously only 1 of 116 evolved mutations was characterized as putatively deleterious (Buskirk et al., 2017). We show that, in contrast to haploid genomes, evolved autodiploid genomes harbor an abundance of putative recessive lethal mutations. We sporulated autodiploids with normal 2N karyotypes by complementing the *MAT $\alpha$*  information on a plasmid. We find evidence of the accumulation of both lethal and deleterious mutations as indicated by

a large number of inviable and slow-growing haploid spores. Autodiploids are initially buffered from the effects of de novo recessive deleterious alleles due to the presence of a second, functional allele. With each successive heterozygous recessive deleterious mutation that fixes, the reduction of functional ohnologs to one eliminates genetic redundancy. Loss of redundancy shifts the distribution of fitness effects (DFE) and an increase in the target size for lethal or deleterious mutations. Over evolutionary time the collective shift in the DFE would impact rate of adaptation.

Interestingly, loss of redundancy occurred rapidly following the historical yeast WGD (Scannell, Byrne, Gordon, Wong, & Wolfe, 2006). Here we show that recessive deleterious and lethal mutations can accumulate shortly after WGD. On a population level, the increased target size for mutations as well as the masking of deleterious mutations may increase standing variation between selective sweeps and may explain populations with deeply diverging clones.

Whole genome duplications occur via autoduplication, wherein the two genomes arise from the same species, or alloduplication, wherein two divergent genomes are brought together through a hybridization event (Madlung, 2013). The WGD events observed here are autoduplications analogous to the origin of autopolyploid taxa (Parisod, Holderegger, & Brochmann, 2010) and to endoreplication events in somatic eukaryotic cells (Fox & Duronio, 2013). The patterns reported here nonetheless inform our understanding of post WGD adaptation. The ancient WGD in the *Saccharomyces* lineage is thought to have occurred by alloduplication followed by LOH at the mating-type locus to restore fertility (Marcet-Houben & Gabaldón, 2015; Wolfe, 2015), and therefore would have gone through an intermediate asexual ‘duplicated’ diploid state, similar to the

*MATa/a* and *MAT $\alpha$ / $\alpha$*  populations investigated here. We demonstrate that this cell type has a direct fitness advantage over an isogenic haploid cell type. The immediate fitness gain of WGD is accompanied by several evolutionary tradeoffs that impact future adaptability including a reduced rate of adaptation, shifted distribution of beneficial mutations, karyotype changes, and the accumulation of recessive deleterious and lethal mutations that reduce redundancy in the duplicated genome.

## Methods

### Strain construction

*MATa/a* strains were constructed for fitness assays by converting yGIL701, a fluorescently labeled *MATa/a* diploid isogenic to our ancestral haploid background, to *MATa/a*. yGIL701 was struck out and 10 separate clones were selected. Clones were transformed with pGIL088, which encodes a gal-inducible *HO* and a *MATa* specific *HIS3* marker. 5 ml cultures of YPD were inoculated with a single transformant for each starting clone. Cultures were grown for 48 hours, allowing for glucose to be depleted and catabolite repression of *GAL* genes to be lifted. After 48 hours 100  $\mu$ l of each culture was plated to SD –his. Histidine prototrophs were screened in  $\alpha$ -Factor (Sigma) for shmoos. Confirmed strains were used in competition assays.

### Evolution experiment

Experimental populations were founded with 130  $\mu$ l of isogenic W303 ancestral culture; 22 with yGIL432 (*MATa*, *ade2-1*, *CAN1*, *his3-11*, *leu2-3,112*, *trp1-1*, *URA3*, *bar1 $\Delta$ ::ADE2*, *hml $\alpha$  $\Delta$ ::LEU2*, *GPA1::NatMX*, *ura3 $\Delta$ ::PFUS1-yEVenus*), and 24 with yGIL646, a *MAT $\alpha$*  strain otherwise isogenic to yGIL432. Populations analyzed here were evolved in separate wells of a 96-well plate. Ancestral strains were grown as 5 ml

overnight cultures from single colonies prior to 96 well plate inoculation. This founding plate was propagated forward and then immediately frozen down.

All populations analyzed here were evolved in rich glucose (YPD) medium. Cultures were grown in unshaken 96-well plates at 30°C and were propagated every 24 hours via serial dilutions of 1:1024. Approximately every 60 generations, populations were cryogenically archived in 15% glycerol.

### **Fitness assays**

Fitness assays were performed as described previously (Buskirk et al. 2017). Evolved autodiploid populations were mixed 1:1 with a version of the ancestral strain (yGIL432 or yGIL646, genotypes listed above) labeled with ymCitrine at *URA3*. Cultures were propagated in a 96-well plate in an identical fashion to the evolution experiment for 40 generations. Every 10 generations, saturated cultures were sampled for flow cytometry. Analysis of flow cytometry data was done using FlowJo 10.3. Selective coefficient was calculated as the slope of the change in the natural log ratio between query and reference strains. Assays were performed for all 46 evolved populations at 16 time points between generations 0 and 4,000.

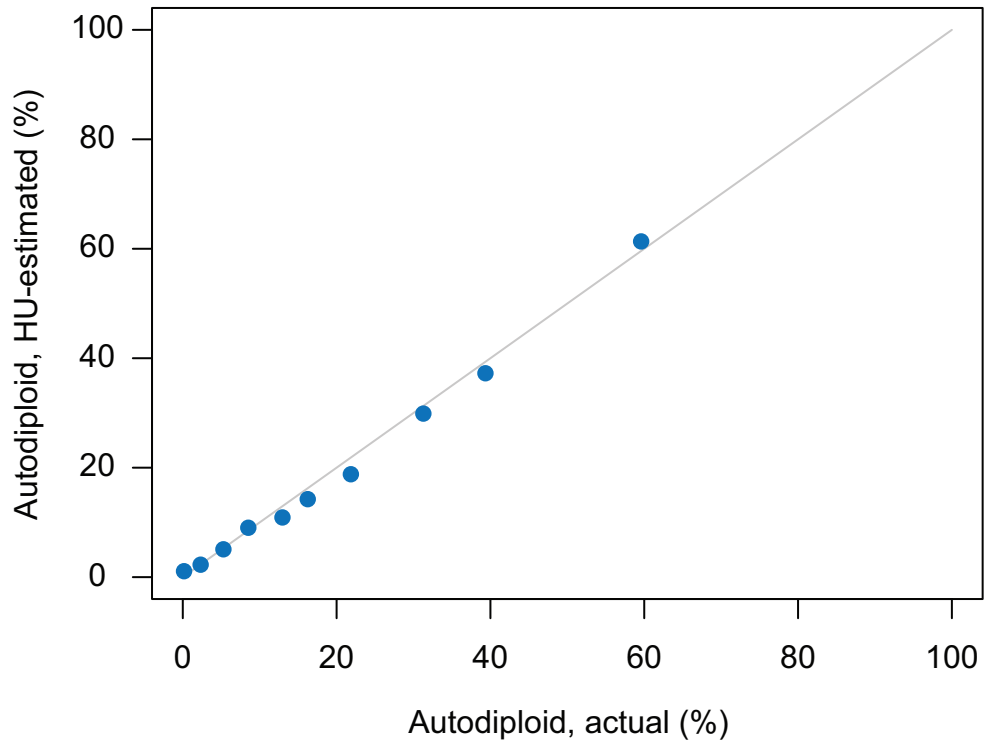
To measure the fitness effect of autodiploidy, fitness assays were performed as described above, using instead a non-labeled version of yGIL432 as a reference. This strain was mixed 1:1 with either a fluorescently-labeled version of the same strain or one of ten biological replicate fluorescently labeled diploid strains. The fitness of each autodiploid reconstruction was calculated as the mean fitness across 12 replicate competitions.

Adaptation rates for each autodiploidized lineage were calculated as the rate of change in relative fitness between generation 0 and the time point at which diploids were present at over 98%. For comparison, rate of adaptation was also calculated for diploid-founded populations evolved in parallel (Marad et al., 2018). The median time point of autodiploid fixation was generation 600 for the haploid-founded dataset. To generate a comparable dataset, rates of adaptation for diploids were calculated from generations 0-600 and 600-4000. Rates were compared in SPSS using a repeated measures ANOVA with two within subject factors (time) and two between subject factors (haploid-founded and diploid-founded). Because some groups violated homogeneity assumptions, post-hoc analysis was done using a Bonferroni correction.

### **DNA content analysis**

Focal populations for DNA content analysis were objectively chosen by randomly selecting one 8-well column per mating-type from the 96-well plate. Time-course ploidy states of 16 focal evolved populations were assayed through flow cytometry analysis of DNA content as described in Gerstein & Otto (2011). Briefly, 10  $\mu$ l of each sample were inoculated in 3 ml YPD and grown overnight. 100  $\mu$ l of saturated cultures were then diluted 1:50 into YPD and grown to mid-log. To arrest in G1, 1 ml mid-log culture was transferred into 200  $\mu$ l 1M hydroxyurea and incubated on a 30°C roller drum for 3 hours. Cultures were then fixed with 70% ethanol, treated with RNase and proteinase K, stained with Cytox green (Molecular Probes), and analyzed on a BD FACSCanto. Haploid and diploid frequencies were estimated using FlowJo v10.3 by fitting data to Watson-Pragmatic cell cycle models. This method of estimation was validated with a series of known ploidy mixtures (**Fig. 2-10**).





**Figure 2-10.** The hydroxyurea (HU) arrest assay and data analysis approach was validated by performing FACS analysis on prefixed control cultures. Measures for ploidy frequency using the assay and analysis were largely accurate when compared to actual measured frequencies. spectrum available to autodiploids.

## Simulations

Simulations of lineage trajectories were performed using a forward-time algorithm designed to imitate the conditions in the evolution experiment reported here. Simulation code, which is described in (Frenkel, Good, & Desai, 2014), was provided by E. M. Frenkel and can be accessed at <https://github.com/genya/asexual-lineage-adaptation>. Estimates for the distribution of fitness effects (an exponential distribution with mean  $s = 0.85\%$ ) and beneficial mutation rate ( $\mu_b = 1.0 \times 10^{-4}$ ) were kept as described previously (Frenkel et al. 2014). This model assumes the spectrum of mutations available to haploids is the same as the spectrum available to autodiploids. Simulations were performed with constant inputs for DFE parameters, beneficial mutation rate, inoculation time of the focal lineage (generation  $t = 0$ ), and fitness advantage of the focal lineage ( $s_0 = 3.6\%$ ). The initial frequency of the focal lineage was varied ( $f_0 = 0.01\% - 1.0\%$ ) for each set of simulations, and a total of ten thousand simulations were performed for each  $f_0$ .

## Sequencing

Evolved clones were obtained by streaking evolved populations to singles on YPD and selecting two clones per population. These clones were grown to saturation in 5 ml YPD and then spun down to cell pellets and frozen at  $-20^\circ\text{C}$ . Genomic DNA was harvested from frozen pellets via phenol-chloroform extraction and precipitated in ethanol. Total genomic DNA was used in a Nextera library preparation. The Nextera protocol was followed as described previously (Buskirk et al., 2017). All individually barcoded clones were pooled and sequenced on 2 lanes of an Illumina HiSeq 2500 sequencer by the Sequencing Core Facility at the Lewis-Sigler Institute for Integrative Genomics at Princeton.

## Sequencing analysis

Two lanes of raw sequence data were concatenated and then demultiplexed using a custom python script (barcodesplitter.py) from L. Parsons (Princeton University). Adapter sequences were trimmed using the fastx\_clipper from the FASTX Toolkit. Trimmed reads were aligned to an S288c reference genome version R64-2-1 using BWA v0.7.12 and variants were called using FreeBayes v0.9.21-24-381 g840b412 (Engel & Cherry, 2013; Garrison & Marth, 2012). Roughly 10,000 polymorphisms were detected between our ancestral W303 background and the S288c reference, and the corresponding genomic positions were removed from analysis. All remaining calls were confirmed manually by viewing BAM files in IGV (Thorvaldsdóttir, Robinson, & Mesirov, 2013). Zygosity was determined based on read depth and allele frequency (Fig. S2B). Mutations were classified as fixed if present in all clones from a population. Clones were genotyped for MAT alleles by identifying mating-type specific sequences within the demultiplexed FASTQ files. Ancestral polymorphisms were inferred using VCFTools (Danecek et al., 2011) to identify homozygosities shared by all clones of the same mating-type. Six mating-type specific SNPs were removed from downstream analysis following verification of homozygosity.

Clone genomes were each independently queried for structural variants. Following BWA alignment, coverage at each position across the genome was calculated. Aneuploidies were detected by calculating median chromosome coverage and dividing this by median genome-wide coverage for each chromosome, producing an approximate chromosome copy number relative to the duplicated genome. CNVs were detected by visual inspection of chromosome coverage plots created in R.

## **Phylogenetic analysis**

Variants identified by SNPEff were used to infer a phylogeny based on 7,932 sites containing 4,742 variable sites, either SNPs or small indels (Fig. S8). Evolved and ancestral sequences (n=93) were aligned with MUSCLE. A general time reversible substitution model with uniform rates (-lnL= 44803.45) was selected based on jModelTest. A maximum likelihood tree was then constructed and rooted by the ancestor in MEGA. Subclades were found to be due to incomplete lineage sorting of mitochondrial polymorphisms. After phylogenetic analysis it was evident that four clones were originally attributed to incorrect populations. Tight clustering and short branch lengths suggests either very recent contamination or an issue during colony isolation (populations were struck out two to a plate on bisected YPD plates). In the text, these clones are identified by the suffix “c” and are attributed to the population to which they are most phylogenetically similar.

## **Identification of common targets and ploidy-enriched targets**

A recurrence approach was utilized to identify common targets of selection (See **Appendix B**). A random distribution of the 3,431 coding sequence (CDS) mutations across all 5,800 genes predicts only two genes to be mutated more than five times by chance alone. We determined the probability that chance alone explains the observed number of mutations of each gene by assuming a random distribution of the 3,431 mutations across the 8,527,393 bp genome-wide CDS. Common targets of selection were defined as genes with five or more CDS mutations and a corresponding probability of less than 0.1%. Notably, analysis using only nonsynonymous mutations identified largely the same set of common targets of selection as did analysis using all CDS mutations. To

determine which targets of selection are impacted by ploidy, our recurrence approach was used to analyze mutations in a previously published *MATa* haploid dataset (Lang et al. 2013; Buskirk et al., 2017). We compared the probability of the observed number of CDS mutations in each gene between ploidies. A gene was considered ploidy-enriched if the ratio of probabilities was at least  $10^5$ .

### **Evolved clone sporulation and tetrad dissection**

Three clones (A02a, B01a, C03b) for which genome sequence data revealed no aneuploidies were selected for sporulation. Evolved *MATa/a* clones were transformed with pGIL071 which encodes the  $\alpha 2$  gene necessary for sporulation and a *URA3* marker for selection. Transformants were sporulated in Spo<sup>++</sup> -ura media. Following 72 hours, sporulation efficiency was calculated via hemocytometer, cultures were digested with zymolyase, and tetrads were dissected on YPD agar plates. Spores were incubated 48 hours and then assayed for germination. Control strain yGIL1039, made by crossing yGIL432 to yGIL646 and converting the resulting diploid to *MATa/a* as described above, was transformed and dissected in parallel.

### **Data Deposition**

The short-read sequencing data reported in this paper have been deposited in the NCBI BioProject database (accession no. PRJNA422100).

## Chapter 3

# Constraints imposed by dominance underlie parallelism in experimental evolution.

*Note - This chapter has been written as a manuscript that we intend to submit within the year.*

### **Abstract**

Factors that limit the traversable sequence space for evolving genes, such as substitution bias, epistasis, and pleiotropy, are broadly referred to as genetic constraints. In an evolving diploid genome the accessibility of beneficial genotypes is additionally constrained by the dominance of underlying mutations. We recently reported an evolution experiment using diploid yeast populations in which some genes under selection show evidence of genetic constraint. Here, we used gene deletions and evolved mutation construction to explore how dominance influences the mutations accumulated in one such gene, *STE4*. We find that complex patterns of dominance, including both underdominance and overdominance, constrain sequence evolution at *STE4*. Furthermore, we show that overdominance can in turn constrain adaptive loss-of-heterozygosity at linked loci.

### **Introduction**

Convergent molecular adaptation is emerging as a frequent pattern observed in comparative genomics (Foote et al., 2015; Hu et al., 2017; Yeaman et al., 2016).

Convergence (or parallelism – depending on the relatedness of the taxa being compared) can be produced by genetic constraints that influence the accessibility of mutational trajectories. Sources of constraint can include limited underlying loci where adaptive mutations can arise (Arendt & Reznick, 2008; Chevin, 2013), factors that mediate the fitness effects of new mutations, and limits to the accessibility of beneficial mutations (Connallon & Hall, 2018).

Constraints operating on the sequence evolution of adapting genes are beginning to be well understood. Empirical work measuring the functionality of possible intermediates between ancestral and evolved protein-coding sequences has shown that, in most cases, only a subset of all possible mutational paths are accessible and the rest are deemed inaccessible because they require an intermediate that is deleterious or unstable (Bridgham, Ortlund, & Thornton, 2009; Gong, Suchard, & Bloom, 2013; Lunzer, Golding, & Dean, 2010; Shah, McCandlish, & Plotkin, 2015; Weinreich, Delaney, Depristo, & Hartl, 2006). These mutation order constraints are largely due to intramolecular epistatic interactions that modify the effects of new mutations (Starr & Thornton, 2016). Similarly, comparative genomic work has shown that pleiotropy may constrain protein sequence evolution, shunting adaptive mutations repeatedly to *cis*-regulatory elements (Wray, 2007).

Dominance is a source of constraint on molecular evolution in diploid organisms that has received little experimental attention. The degree of dominance of *de novo* mutations has the potential to constrain mutational trajectories in a similar manner to epistasis by making some routes improbable or even selectively inaccessible. The probability of a given beneficial mutation fixing in a population is theoretically

determined as a product of its coefficient of selection and its dominance; therefore dominant beneficial alleles have a greater probability of fixation than recessive ones (Haldane, 1924). This constrains evolution by making recessive beneficial mutations less accessible, a phenomenon coined “Haldane’s sieve”.

Attenuated access to beneficial mutations due to low degrees of dominance has an overt affect when comparing haploid and diploid asexual yeast adaptation in the laboratory. Haploid populations adapt more quickly than diploids (Fisher et al., 2018; Gerstein et al., 2011; Marad et al., 2018) and haploids and diploids differ in the identities of beneficial mutations (Fisher et al., 2018). The constraints of recessive or partially beneficial mutations can be overcome in asexual populations through loss-of-heterozygosity (LOH) whereby a beneficial mutation becomes homozygous as a result of mitotic recombination (Gerstein et al., 2014; Smukowski Heil et al., 2017). However, the likelihood of overcoming Haldane’s sieve is also constrained by rates of LOH, which vary markedly in frequency across the genome (Fisher et al., 2018).

Certain types of dominance change the sign of a mutation’s fitness effect depending on its zygosity – namely underdominance and overdominance (when the heterozygote is the least or most fit genotype, respectively). In asexual populations underdominance should constrain locus-specific adaptation by impeding access to a potentially adaptive homozygous genotype. Inversely, overdominant mutations should be able to establish in populations and promote the maintenance of variation (Fisher, 1928). The frequency and significance of overdominant mutations, however, remains a matter of debate. While genome scans have turned up little evidence of overdominance (Goudie, Allsopp, & Oldroyd, 2014; Hedrick, 2012; Szulkin, Bierne, & David, 2010), recent



theory suggests overdominance may be a frequent, if transient, outcome of diploid evolution (Manna, Martin, & Lenormand, 2011; Sellis et al., 2011). Studies of laboratory-evolved populations have found evidence of overdominant effects of both *de novo* mutations (Sellis, Kvitek, Dunn, Sherlock, & Petrov, 2016) and standing variation (Chelo & Teotónio, 2013) during short-term adaptation in diploid populations, however it remains unclear whether these are rare examples or represent a common mode of adaptation.

Experimental evolution tests evolutionary hypotheses by determining the distribution of evolutionary outcomes across a large numbers of replicate populations. It is then possible to interrogate sources of constraint on sequence evolution by testing the fitness consequences of genetic routes not observed. In other words, we can examine why certain evolutionary paths were not realized. We recently reported 20 genes mutated with a significant degree of parallelism across 46 populations of 4,000-generation laboratory-evolved autodiploid yeast (Fisher et al., 2018). Among the genes identified as parallel targets of selection is *STE4*, which encodes the highly conserved beta subunit of the heterotrimeric G protein complex along with *STE18* and *GPA1* (Whiteway et al., 1989). In yeast, binding of mating pheromone to the Ste2 (or Ste3) receptor releases Ste4/Ste18, which activate a MAP-kinase cascade ultimately eliciting a cell-cycle arrest and a transcriptional response. In the absence of pheromone binding, the mating pathway is on at a basal level. For asexually dividing cells, elimination of costly basal signaling through the mating pathway has been shown to have a fitness benefit (Lang et al., 2009), and correspondingly *STE4* loss of function has previously been identified as adaptive in evolving haploid yeast populations (Lang et al., 2013). However, in autodiploids –

diploid yeast that are homozygous at the mating-type locus and thus mating-competent – evolved mutations in *STE4* are heterozygous and they cluster in a small region of the coding sequence.

We hypothesized that the discrepancy between patterns of sequence evolution in haploids and autodiploids is due to constraints imposed by dominance. Here, we use *STE4* as a case study to explore how dominance constrains sequence evolution and produces patterns of parallelism in the locations of evolved mutations. We find complex dynamics of dominance governing *STE4*, with both underdominance and overdominance operating to constrain sequence evolution. We then investigate how constraints operating at a single locus can affect gross chromosomal sequence evolution. We find that overdominance at a single locus can restrict LOH of linked adaptive mutations.

## **Methods**

### **Evolved alleles**

Evolved *STE4* mutations were identified in sequencing data reported in Fisher et al. (2018). The probability that all mutations occurred in the observed region was calculated using a one-sample proportions test.

### **Construction of evolved mutation and *STE4* deletion strains**

We constructed strains in order to assay the effects of complete gene deletions and evolved *STE4* mutations on fitness (**Appendix B, table B-1**). All strains were constructed in the same W303 ancestral background (yGIL121; *MATa*, *URA3*, *ade2-1*, *his3-11,15*, *leu2-3,112*, *trp1-1*, *CAN1*, *bar1Δ::ADE2*, *hmlαΔ::LEU2*). Briefly, deletion strains were generated by integrating the *ste4Δ::KanMX* locus from the deletion collection. Evolved mutants were generated via CRISPR/Cas9 allele swaps as described

in (Fisher, Kryazhimskiy & Lang et al., 2019). A synonymous mutation, Thr326Thr, was introduced along with each evolved mutation to ablate the gRNA recognition site, and mutants carrying this mutation in isolation were assayed alongside double mutants to verify neutrality. Because *ste4* mutations arose in the context of autodiploid evolution, all diploid genotypes were converted to *MATa/a* by transforming *MATa/α* diploids with pGIL088, which contains a galactose-inducible *HO* homing endonuclease, as reported in (Fisher et al., 2018). For fitness assays and cytometry analysis, eight replicate *MATa/a* colonies were picked for each mating-type conversion. Full details on strain construction are provided in Appendix B.

### **Fitness assays**

We measured the effects of complete gene deletions and evolved *STE4* mutations on fitness using competitive fitness assays as previously reported (Buskirk et al., 2017). Briefly, query cultures were mixed 1:1 with a ploidy and *MAT* genotype-matched fluorescently labeled ancestral strain. Co-cultures in a 96-well plate were propagated for 50 generations in a manner identical to the evolution experiment in the variants arose. Saturated cultures were sampled for flow-cytometry at 10-generation intervals. Flow cytometry data were analyzed with FlowJo 10.3. Selective coefficients were calculated as the slope of the best-fit line of the natural log of the ratio between query and reference strains against time using custom R scripts.

Two technical replicates of eight biological replicates were averaged for analysis of all *MATa/a* genotypes and all deletion mutants. Evolved mutations in a haploid background were estimated from four replicates of a single correct clone. Fitness data for haploid and diploid genotypes were analyzed independently using a one-way analysis of

variance (ANOVA). Post hoc comparisons using the Tukey test were carried out to identify genotypes with significantly different fitness than wild type controls.

### **Short-term evolution experiment**

We investigated the effect of evolved *ste4* alleles on likelihood of loss-of-heterozygosity (LOH) at linked locus, *WHI2*. We first validated the fitness benefits of homozygotes and heterozygotes for an evolved *whi2*C85T (Q29\*) allele via mutant reconstruction and fitness assays as described above. We then generated strains containing dominant drug cassettes tightly linked to the *WHI2* locus to investigate the effect of *STE4* linkage on loss of heterozygosity along the right arm of Chr. XV (**Appendix B – supplementary methods**).

Three strains (*WHI2*::HphMX-*STE4*/*WHI2*::KanMX-*STE4*, *WHI2*::Hph-*STE4*/*whi2*Q29\*::KanMX-*STE4*, and *WHI2*::Hph-*STE4*/*whi2*Q29\*::KanMX-*ste4*Glu315\*) were grown in 10ml overnight cultures in YPD +2x G418 & 2x Hygromycin. Saturated cultures were diluted 1:1,000 to initiate 96 128 µl cultures across three 96-well plates. 96-well plates were incubated unshaken at 30°C and propagated daily in an identical fashion to the original evolution experiment in which the mutations arose (Fisher et al. 2018). After 500 generations of evolution heterozygosity was assayed by automated spotting of a 2µl volume containing ~5,000 cells per population to 2x double drug and 2x single drug YPD agar plates. Plates were inspected for speckled spots (indicating polymorphisms) and absence of growth (indicating sweeps of homozygous genotypes). We compared the number of populations with evidence of LOH polymorphism or sweeps between genotypes using a Fisher's exact test with a Bonferroni correction for multiple comparisons.

## Analyses

All statistical analyses reported were performed using tools in the R Stats package in R v. 3.4.0. Data plots were produced in R using the ggplot2 package (Wickham, 2016).

## Results

A one-sample proportions test shows that mutations accumulated nonrandomly across the linear 1,276 bp sequence of the *STE4* gene, ( $X^2(1, N=6) = 18.76, p < 10^{-4}$ ). Six evolved mutations cluster in a 260 base pair window that encompasses only 20% of the coding sequence (**figure 3-1a, table 3-1**). Three mutations likely result in a truncation of the C-terminus. Two missense mutations were observed, Ala287Ser and Arg312Gln, along with one synonymous mutation.

### ***STE4* loss-of-function is underdominant**

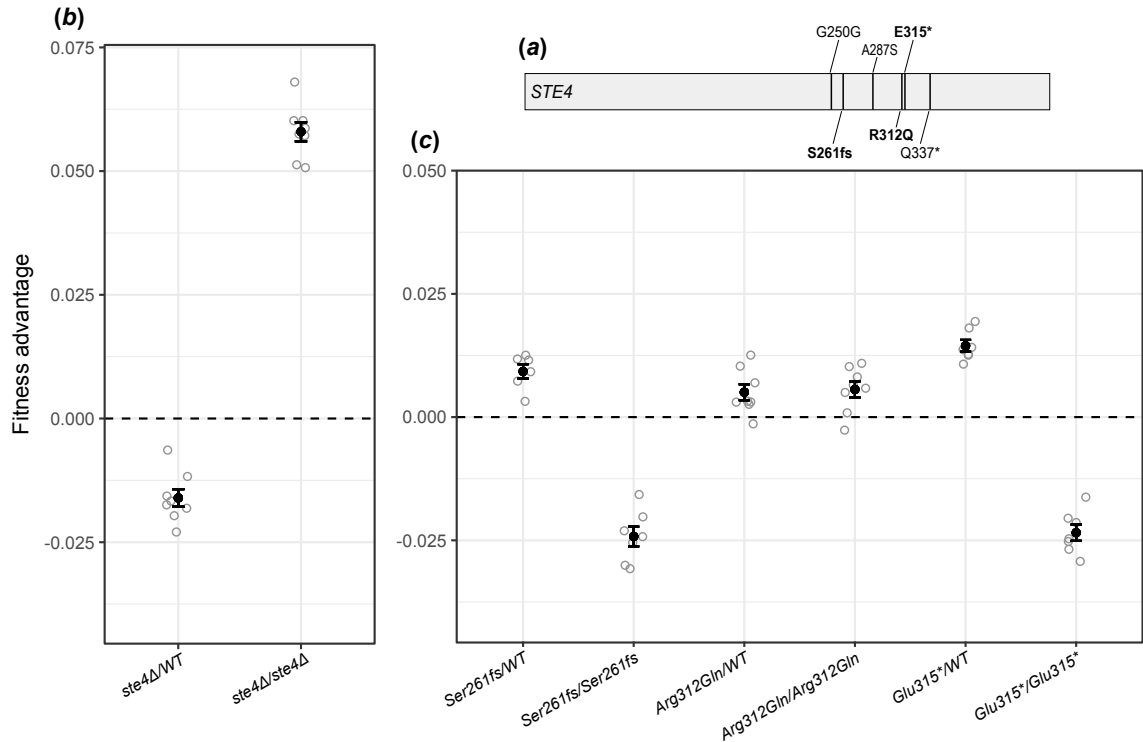
We first tested the extent to which complete loss of function of one or both copies of *STE4* impacts fitness. We generated *STE4* deletion (*ste4Δ*) strains as haploids, as heterozygous *MATa/a* diploids, and as homozygous *MATa/a* diploids. As reported previously we find that *ste4Δ* is beneficial in a haploid (**figure 3-2**). Similarly, homozygous deletion mutants are beneficial in *MATa/a* diploids. Surprisingly, however, heterozygous *ste4* deletion mutants are substantially less fit than wild type (**figure 3-1b**). *STE4* loss-of-function (LOF) alleles, therefore, are underdominant — less fit than either homozygous wild type or homozygous LOF.

### **Evolved *STE4* alleles are overdominant**

We next examined the fitness effects of three evolved *ste4* alleles: one frameshift, one nonsense, and one missense mutation (**figure 3-1a, table 3-1**). Alleles were assayed in haploids and in both homozygous and heterozygous state in *MATa/a* diploids. Because

there is an effect of ploidy on fitness (Fisher et al., 2018), haploid and diploid fitness data were analyzed separately. A one-way ANOVA confirmed a significant interaction of genotype and fitness amongst haploid genotypes ( $F(5, 26)=21.2, p < 10^{-3}$ ), and diploid genotypes ( $F(8, 47)=73.63, p < 10^{-3}$ ). Post-hoc comparisons were used to identify non-neutral genotypes that differed significantly from wild type. Alleles containing only the synonymous Thr326Thr substitution, introduced as part of Cas9-mediated allele swaps (see methods), did not significantly differ from wild type in any context (**figure 3-3**).

In a haploid background the evolved frameshift and nonsense alleles had a substantial fitness benefit ( $p < 10^{-4}$  both genotypes) that surpassed the deletion ( $p=0.046, p=0.026$ , respectively) while the missense allele appears neutral ( $p=0.995$ ) (**figure 3-2, table 3-1**). Because all evolved mutations are maintained as heterozygous in the evolution experiment in which they were identified, we predicted that the fitness effects of evolved alleles would exhibit some degree of dominance. Indeed, heterozygous evolved alleles showed positive dominance and conferred ~40-50% of the haploid fitness advantage (**figure 1b, table 3-1**). Again, the frameshift and nonsense mutations appear beneficial ( $p=0.02, p < 10^{-4}$ , respectively) while the missense appears neutral ( $p=0.49$ ). Fitness advantages of homozygous Ser261fs and Glu315\* mutants are significantly different than that of heterozygous mutants ( $p < 10^{-3}$  for both comparisons). However, rather than showing an additive fitness effect, homozygous mutants of both mutations have a strong fitness defect of ~2.4%. Evolved *ste4* alleles therefore are overdominant — most beneficial as a heterozygote.



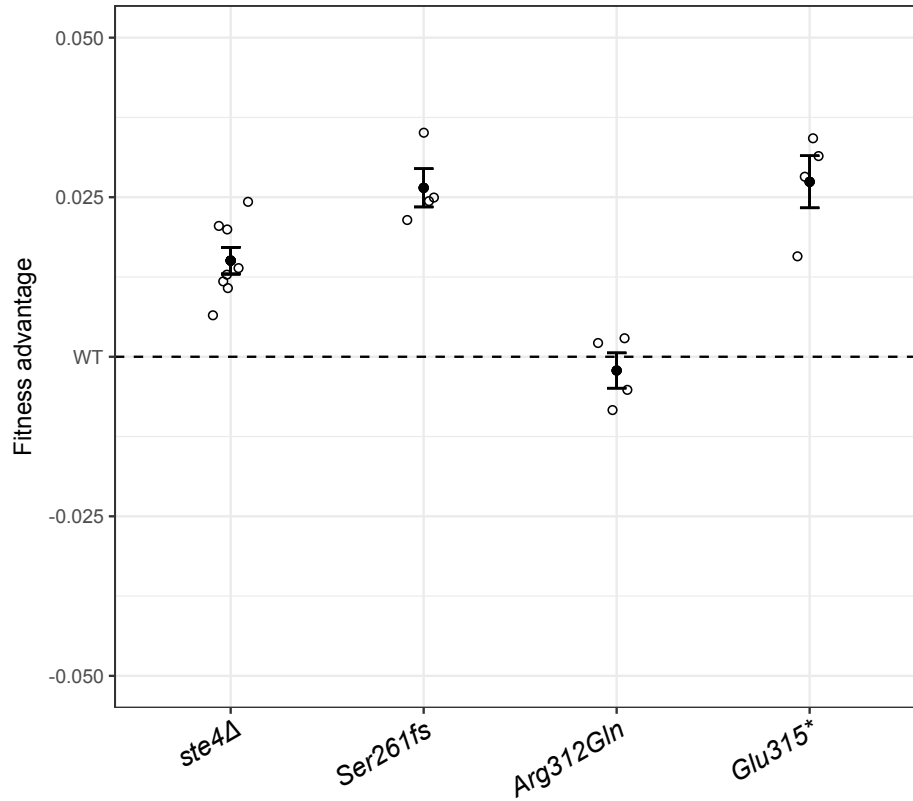
**Figure 3-1.** *a)* The positions of 6 mutations that independently arose in *STE4* in Fisher et al. (2018) positioned along the coding sequence. Bolded mutations were reconstructed for fitness assays. *b)* Heterozygous deletion (*ste4Δ/WT*) clones exhibited a decrease in relative fitness while homozygous deletions (*ste4Δ/ste4Δ*) have a large fitness benefit. *c)* The evolved frameshift and nonsense alleles have slight fitness benefits as heterozygotes and are strongly deleterious when homozygous. The evolved missense mutation appears neutral. Open points in *b-c* represent selection coefficients from eight technical replicates. Bold point is the mean  $\pm$  standard error.

**Table 3-1.** Evolved *STE4* mutations.

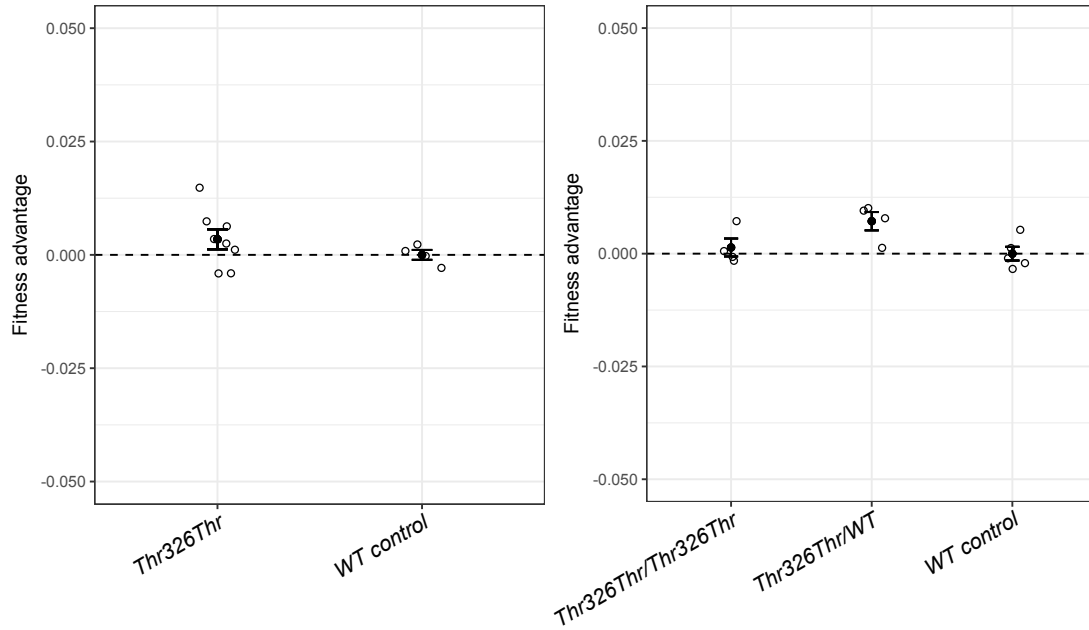
Nuclear mutation	Protein effect	Fitness effect		
		Haploid	Heterozygous	Homozygous
G750C- heterozygous	Gly250Gly	NA	NA	NA
*T781Δ- heterozygous	Ser261fs	+2.6%	+0.92%	-2.4%
G859T- heterozygous	Ala287Ser	NA	NA	NA
*G935A- heterozygous	Arg312Gln	-0.2% (neutral)	+0.5%(neutral)	+0.5%(neutral)
*G943T- heterozygous	Glu315*	+2.7%	+1.4%	-2.3%
C1009T- heterozygous	Gln337*	NA	NA	NA

\*Mutations that were reconstructed as part of this study.





**Figure 3-2.** Fitness effects of *ste4Δ* and evolved mutations in a haploid background. The data confirm a previously reported benefit of the knockout (Lang et al., 2009). Two of the three evolved mutations have a significant fitness benefit while one, the only non-truncation allele, appears neutral. Open points represent selection coefficients from eight technical replicates for *ste4Δ* and four replicates of each evolved genotype. Bold point is the mean  $\pm$  standard error.

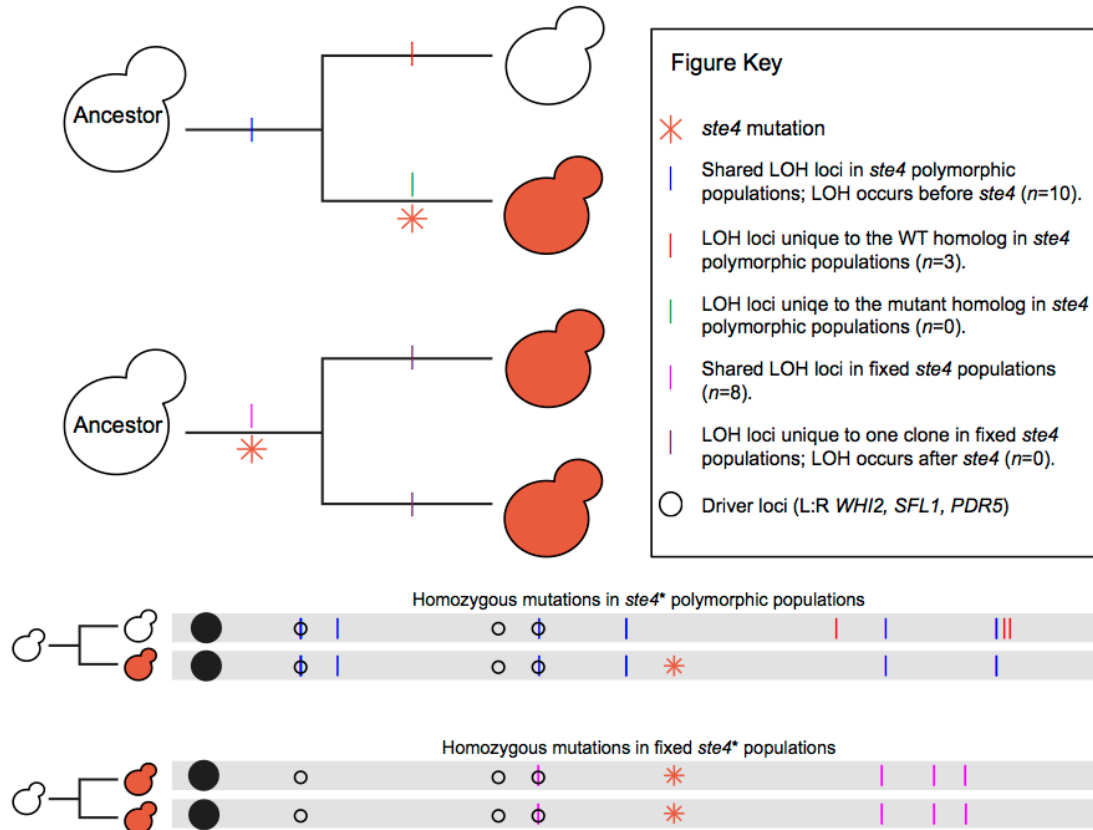


**Figure 3-3.** A synonymous Thr326Thr *STE4* mutation was introduced along with each evolved mutation during Cas9-mediated allele swaps in order to ablate the PAM site of the gRNA target sequence. Single mutants for the synonymous mutation were isolated and assayed along with double mutants carrying evolved mutations. WT Synonymous mutant fitness did not differ from wild type in a haploid background ( $p=0.93$ ), a heterozygous diploid background ( $p=0.24$ ), or a homozygous diploid background ( $p=1.0$ ). Open points represent selection coefficients from eight technical replicates for the mutation in a haploid background and four replicates of each diploid genotype. Bold point is the mean  $\pm$  standard error.

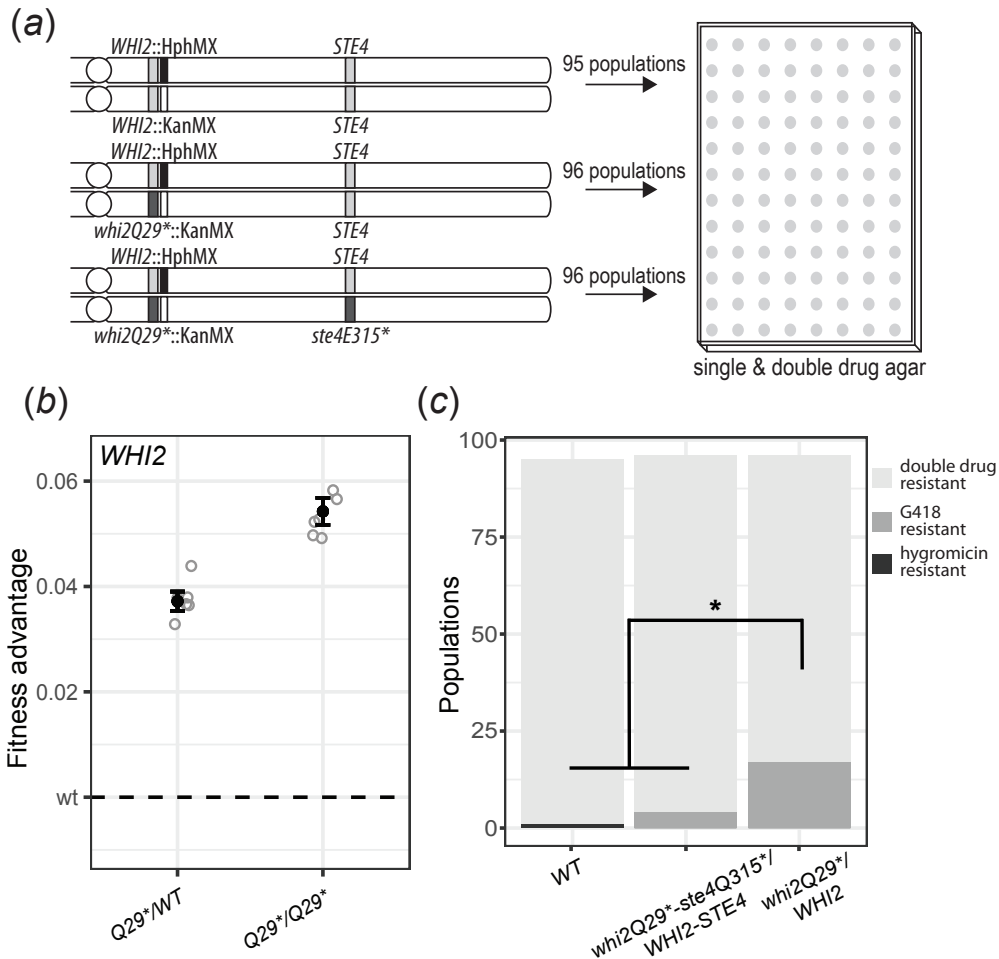
### **Overdominant *STE4* alleles impede adaptive LOH of linked beneficial mutations**

Overdominance of evolved mutations at the *STE4* locus explains why these mutations are only ever observed in heterozygous state. This is initially unsurprising, as these mutations arose during the asexual propagation of autodiploids and most other mutations identified are heterozygous. There are, however, two large genomic regions prone to high rates of loss-of-heterozygosity (LOH) as indicated by the clustering of homozygous mutations (Fisher et al., 2018). The *STE4* locus is contained within one of these regions on the right arm of Chr. XV. Three loci centromeric to *STE4* (*WHI2*, *SFL1*, & *PDR5*) are identified as parallel targets of selection in the preceding evolution experiment and evolved alleles of all three genes are commonly observed as homozygous.

We considered whether the overdominance of evolved *ste4* mutations could constrain adaptive LOH at linked loci. Where we can resolve the timing of events, we find that LOH events do not occur after a *STE4* mutation arises on the right arm of Chr. XV (**figure 3-4**). We hypothesize that once they arise, overdominant *STE4* alleles prevent adaptive LOH at a linked locus, *WHI2*. We chose an evolved nonsense *WHI2* allele to reconstruct and confirmed that LOH of this mutation would be beneficial (**figure 3-5b**). We then conducted a 500-generation evolution experiment to determine whether a *WHI2* mutation is less likely to lose heterozygosity when linked to an evolved *STE4* allele. Heterozygosity was tracked via dominant drug markers tightly linked to the *WHI2* locus (**figure 3-5a**). After 500 generations we find a significantly higher rate of LOH in populations where the adaptive *WHI2* allele is unlinked to an overdominant *STE4* mutation compared to populations carrying a distal *STE4* mutation (17/96 to 4/96,



**Figure 3-4.** Multiple clones were sequenced for most populations reported in Fisher et al. (2018). We used these data to look for indirect evidence of *STE4* mutations influencing rates of loss-of-heterozygosity along the right arm of Chr. XII. In particular, populations polymorphic for evolved mutations were used to look for LOH events that occurred before the emergence of *STE4* mutations, and populations in which evolved *ste4* alleles are fixed were used to look for evidence of LOH events that occurred after *STE4* mutations. There was no evidence of LOH occurring after *STE4* mutations (purple lines), while there was ample evidence of such events preceding *STE4* mutations (blue lines). Driver loci (circles) are those genes identified as common targets of selection in the prior evolution experiment.



**Figure 3-5. a)** Dominant drug markers were tightly linked to the *WHI2* locus to track heterozygosity of wild type and *whi2Q29\** alleles. **b)** Loss of *whi2Q29\** heterozygosity is adaptive. Open points represent selection coefficients from eight technical replicates. Bold point is the mean  $\pm$  standard error. **c)** The number of populations that experienced LOH at *WHI2* as measured by loss of double drug resistance, \* $p=0.013$ .

$p=0.013$ ). Rates of heterozygosity loss in populations with linked *WHI2 STE4* mutations are not different from populations initially wild-type at all loci (4/96 to 1/95,  $p=1$ ) (**figure 3-5c**).

## **Discussion**

This case study finds an unexpected level of mechanistic complexity underlying the constraints of dominance at a single experimentally evolved locus. We report here that the collective effects of underdominance and overdominance direct molecular parallelism at a single adapting locus and that these factors in turn constrain adaptation at linked loci. Constraints on sequence trajectories imposed by dominance within a single gene have not previously been explored in detail, so it is difficult to know whether a ruggedness produced by dominance is a common feature of the fitness landscape of evolving genes. Such ruggedness is, however, similar to studies showing that epistasis restricts accessible routes from ancestral to derived protein sequences (Ferretti, Weinreich, Tajima, & Achaz, 2018).

We show here that underdominance of loss-of-function mutations shunts adaptive mutations to a predictable region of *STE4*. Most theory addressing mutational dominance and constraint focuses on the consequences of recessiveness, namely the constraints imposed by Haldane's sieve (Charlesworth, 1998; Orr & Betancourt, 2001) and the load imposed by recessive deleterious mutations (Charlesworth & Charlesworth, 1999; Chasnov, 2000). Underdominance is most frequently invoked as a cause of reproductive isolation (Barton & De Cara, Maria Angeles Rodriguez, 2009), but our findings suggest an underappreciated role in evolutionary constraint. Underdominant variants are theorized to be able to fix only in extremely small or fragmented populations (Newberry,

McCandlish, & Plotkin, 2016) and thus would efficiently block access to fitness optima of homozygous genotypes. For example, modeling done by Stewart et al. (2013) suggested mutational inaccessibility due to underdominance might underlie why negative autoregulation is a common property of bacterial transcription factors but is rare in diploid organisms. How pervasively restrictive underdominance may be depends on the distribution of dominance among mutations. Comprehensive analyses of the gene deletions in yeast reveal few underdominant deletions (Agrawal & Whitlock, 2011), however, this may not be indicative of the distribution of single nucleotide polymorphism dominance.

The evolved mutations assayed here demonstrate a strong degree of overdominance. Recent theoretical examination of adaptation in diploids has renewed interest in the significance of overdominant mutations in adaptation and suggested overdominant polymorphisms may be a frequent mode of adaptation (Manna et al., 2011; Sellis et al., 2011). These models find that when selection on a trait is stabilizing, strong effect heterozygous mutations that overshoot the fitness optimum as homozygotes should be somewhat common. Experimental evolution remains the best way to test this prediction. Previously, the only examples of overdominance arising *de novo* in laboratory evolution were amplifications of glucose transporter genes in glucose-limited media (Sellis et al., 2016). Overdominance of a copy number variant is well explained by an “overshoot” of an optimal gene copy number. The *STE4* truncations reported here represent the first evidence of experimentally evolved overdominant point mutations. The mechanism(s) underlying underdominance of *STE4* deletions and overdominance of evolved *STE4* alleles are not clear. It is also not clear that the low fitness of the

hemizygote and the low fitness of the homozygous evolve mutations share the same molecular underpinning. Future work is needed to determine if these mutations also push a trait of an optimum maintained by stabilizing selection.

The experiments reported here begin to address the conflict between two modes of adaptation in asexual diploids – overdominant mutations and loss-of-heterozygosity. Mitotic recombination resulting in loss-of-heterozygosity (LOH) is a common mechanism of adaptation in laboratory evolving diploid yeast (Fisher et al., 2018; Gerstein et al., 2014; Smukowski Heil et al., 2017). Despite its prevalence, we do not know the specific mechanism(s) by which LOH occurs in this or previous studies. Most evidence suggests that the repair of DNA lesions that occur in G1 or S phase results in reciprocal crossovers and directional gene conversions, each of which can be mediated by different mechanisms of break repair (Charles & Petes, 2013; Prado, Cortes-Ledesma, Huertas, & Aguilera, 2003). Most reported instances of LOH in asexual yeast adaptation involve a conversion tract that runs from the break point to the telomere. This means that there is effective linkage between loci that are kilobases apart. In this study we report an effect of this linkage and a difference in the rates of LOH depending on the tract length. The difference in rates between conversion tracts that run breakpoint to telomere and shorter conversion tracts is consistent with studies suggesting these occur by different mechanisms and that the former is more frequent than the latter (LaFave & Sekelsky, 2009; Lee et al., 2009). This is the first evolution experiment to directly measure rates of LOH when distal conversion is unfavorable. Our findings indicate adaptation via mitotic recombination can be constrained by distal heterozygosity.



To our knowledge, this is the first functional characterization of how dominance constrains adaptation at a single locus. We find that mutations at *STE4* do not follow a simple spectrum from recessive to dominant. Future work to examine the biological basis of idiosyncratic dominance at *STE4*, as well as how dominance operates at other loci under selection, will shed light on the pervasiveness of the patterns reported here.

## Chapter 4

# Leveraging parallelism to detect genetic interactions between evolved mutations.

*Note- The work described in this chapter will be published in a theme issue on convergent evolution in Philosophical Transactions B<sup>1</sup>. This work was performed in collaboration with Sergey Kryazhimskiy of the Division of Biological Sciences at the University of California, San Diego. As a co-senior author, Sergey co-led the design of the study, the analysis of the data, and the writing of the paper.*

### **Abstract**

Eukaryotic genomes contain thousands of genes organized into complex and interconnected genetic-interaction networks. Most of our understanding of how genetic variation affects these networks comes from quantitative-trait loci (QTL) mapping and from the systematic analysis of double deletion (or knockdown) mutants, primarily in the yeast *Saccharomyces cerevisiae*. Evolve and re-sequence experiments are an alternative approach for identifying novel functional variants and genetic interactions, particularly between non-loss of function mutations. These experiments leverage natural selection to obtain genotypes with functionally important variants and positive genetic interactions.

---

<sup>1</sup> Fisher, K. J., Kryazhimskiy, S., Lang, G.I. (2019). Detecting genetic interactions using parallel evolution in experimental populations. *Philosophical Transactions of the Royal Society B*, 10.1098/rstb.2018.0237

However, no systematic methods for detecting genetic interactions in these data are yet available. Here, we introduce a computational method based on the idea that variants in genes that interact will co-occur in evolved genotypes more often than expected by chance. We apply this method to a previously published yeast experimental evolution data set. We find that genetic targets of selection are distributed non-uniformly among evolved genotypes, indicating that genetic interactions had a significant effect on evolutionary trajectories. We identify individual gene pairs with a statistically significant genetic interaction score. The strongest interaction is between genes *TRK1* and *PHO84*, genes that have not been reported to interact in previous systematic studies. Our work demonstrates that leveraging parallelism in experimental evolution is useful for identifying genetic interactions that have escaped detection by other methods.

## **Introduction**

Determining the extent to which genetic variants interact to affect phenotypes is a central challenge in biology. Traditional methods such as QTL mapping and double deletion analysis have proven useful for identifying functional variants and genetic interactions in laboratory model systems such as the yeast *Saccharomyces cerevisiae*. However, both of these methods have limitations. QTL mapping provides a robust approach to identifying natural genetic variants that contribute to complex traits, but most studies are underpowered to detect genetic interactions. Large studies (with on the order of  $10^3$  segregants) have shown that QTL-QTL interactions contribute to a wide array of complex traits (Bloom, Ehrenreich, Loo, Lite, & Kruglyak, 2013; Bloom et al., 2015; Huang et al., 2012; Wilkening et al., 2014) but even the largest study to date did not have the statistical power to identify small-effect interactions (Bloom et al., 2015). In addition,

genetic linkage makes it difficult in many cases to identify the causal variants underlying most QTLs.

Systematic phenotypic screens of double deletions/knockdowns in yeast and other organisms avoid these problems (Babu et al., 2014; Costanzo et al., 2016; Lehner, Crombie, Tischler, Fortunato, & Fraser, 2006; Tong et al., 2004). These types of studies have successfully identified a large number of genetic interactions, particularly within protein complexes (Baryshnikova et al., 2010). By design, this approach is limited to detecting only strong pairwise interactions between loss-of-function variants. Most natural variation, however, is not loss-of-function (Bergström et al., 2014; Saleheen et al., 2017), and thus a comprehensive picture of genetic interactions will require tests of interactions between functional variants.

An alternative approach to identifying functionally important variants and interactions between them is to leverage the power of natural selection. When different populations of the same or different species face the same environmental challenge, natural selection often finds the same phenotypic (Hagen & Gilbertson, 1972; O'quin et al., 2010; Protas et al., 2006) or even genetic (Glazer et al., 2014; McCracken et al., 2009; Zhen et al., 2012) solution to this challenge. This phenomenon is referred to as convergent or parallel evolution. Thus, the observation of parallel genetic changes in multiple independent lineages can be used to identify variants that contribute to functionally important traits. This approach has been successful in identifying key mutations in pathogen and tumor evolution (Carroll et al., 2015; Gerlinger et al., 2014; Kryazhimskiy, Sergey, Bazykin, Plotkin, & Dushoff, 2008; Lieberman et al., 2011). The idea of convergence or parallelism has also been used to detect epistasis within genes

(Codoñer & Fares, 2008; Korber, Farber, Wolpert, & Lapedes, 1993; Kryazhimskiy, Sergey, Dushoff, Bazykin, & Plotkin, 2011; Lockless & Ranganathan, 1999; Shapiro, Rambaut, Pybus, & Holmes, 2006) and more recently also between genes (Neverov, Kryazhimskiy, Plotkin, & Bazykin, 2015) in natural populations. In this type of analysis, pairs of variants are identified as genetically interacting if they co-occur in the same genotype more often than expected by chance. There are three challenges in using parallelism to detect functional variants and genetic interactions in natural populations. First, true functional parallelism is confounded by common ancestry. Second, because we rarely know what selection pressures drove the evolution of the functional variants, it is difficult to connect genotype with phenotype. Third, detecting epistasis requires many variants to accumulate and is therefore only feasible in either fast evolving populations or over very long time-scales.

Evolve and re-sequence experiments offer a complementary approach for detecting functional variants and genetic interactions. Like inferences from natural populations, this approach also relies on selection to find functional variants and genetic interactions between them. This approach, however, overcomes problems arising in studies of naturally evolving populations. Hundreds of replicate microbial populations can be propagated in identical conditions such that the selected phenotypes are either known or can be measured (Long, Liti, Luptak, & Tenaillon, 2015). After hundreds or thousands of generations, entire populations or individual isolated clones are sequenced, and adaptive variants are identified by their parallel occurrence in replicate lines (e.g. (Barrick et al., 2009; Fisher et al., 2018; Good, McDonald, Barrick, Lenski, & Desai, 2017; Lang et al., 2013; Tenaillon et al., 2012)). Since replicate populations evolve

independently, overabundance of parallel variants is a signal of positive selection, which is not confounded by common ancestry. Genetic interactions are known to contribute to adaptive evolution (Phillips, 2008), and the data from evolve and re-sequence experiments must contain information about these genetic interactions. To the best of our knowledge, only one study so far has leveraged this type of data to detect epistasis and demonstrate how it affected evolutionary trajectories (Tenaillon et al., 2012). The challenge is that large data sets are required to detect overrepresented pairs of genes that contain interacting variants. However, unlike in QTL mapping approaches, the number of variants in experimentally evolved populations can be controlled to increase statistical power to reveal genetic interactions. At the same time, evolution in the lab, just like evolution in nature, assesses all types of variants, which in principle allows us to detect genetic interactions that may not be revealed in gene-deletion studies.

Here, we present an approach that leverages parallelism in experimental evolution to detect genetic interactions between genes that acquire mutations independently across populations. We detect genetic interactions between pairs of genes using mutual information (Bindewald & Shapiro, 2006; Gloor, Martin, Wahl, & Dunn, 2005; Kim, Koyutürk, Topkara, Grama, & Subramaniam, 2005). This quantity captures the statistical dependence between the occurrences of mutations at two specific loci in the same genotype. We use this approach to analyze a recently published whole-genome dataset derived from experimentally evolved asexual populations of yeast. We find that the accumulated mutations are distributed between genotypes non-uniformly, indicating that genetic interactions have contributed to adaptive evolution in these laboratory populations. We identify specific pairs of genes that have acquired mutations in parallel

more often than expected by chance, indicating putative genetic interactions. We experimentally verify that our top-hit pair, *TRK1* and *PHO84*, shows a positive genetic interaction when reconstructed in the ancestral background.

## **Materials and Methods**

### **Sequencing data re-analysis**

Evolved mutations used for this analysis were obtained from 92 endpoint clones isolated from 42 populations of 4,000 generation evolved autodiploids, previously reported in Fisher et al. (2018). Populations were grown in rich media in individual wells of unshaken 96-well plates at 30°C and diluted 1:1024 every 24 hours. At approximately 60 generation intervals populations were cryoarchived in 15% glycerol. We reanalyzed the raw sequencing data to improve annotation quality. All raw data files were demultiplexed using a custom python script (`barcodesplitter.py`) from L. Parsons (Princeton University). Adapter sequences were trimmed using `fastx_clipper` (FASTX Toolkit). Reads were then aligned to a customized W303 genome using BWA v0.7.12 (Li & Durbin, 2009). VCFtools was used to filter variants common to all samples and mating-type specific polymorphisms (see Fisher et al., 2018)). Remaining polymorphisms were then annotated using a strain-background customized annotation file (Matheson et al., 2017).

### **Calculating mutual information**

We used the evolved mutations generated by reprocessed sequence data to look for evidence of genetic interactions. To prevent false positives due to common ancestry, only one clone with the most mutations from each population was included in the analysis. We then excluded all intergenic and synonymous mutations. Lastly, to reduce

the number of statistical tests, we looked for genetic interactions only among “multi-hit” genes, i.e., those in which at least three mutations in independent populations were detected in the data set. This was done to reduce noise by enriching for beneficial mutations. Nevertheless, we estimate, by simulation controlling for gene length, that 12% of genes receive three or more mutations by chance alone and are likely neutral. This reduced data set includes 113 “multi-hit” genes from 46 independently evolved genotypes.

For all 6,328 pairwise combinations of multi-hit genes we calculated the mutual information statistic. To do so, we model an evolved genotype with a series of (possibly non-independent) Bernoulli random variables  $\sigma_i$  with  $i = 1, 2, \dots, K$ , where  $K = 113$ , the total number of genes where mutations can possibly occur.  $\sigma_i$  takes value 1 if a mutation occurs (in the data) in gene  $i$  and it takes values 0 if it does not occur. We first estimate the marginal probability of a mutation occurring in gene  $i$  as

$$(3) \quad P(\sigma_i = 1) = C_M \sum_{g=1}^N \tilde{M}_{gi}.$$

Here,  $\tilde{M}_{gi} = M_{gi} + \varepsilon$  and  $M_{gi} = 1$  if mutation in gene  $i$  is present in genotype  $g$  in the data. We regularize our estimates by adding a pseudocount  $\varepsilon = \frac{1}{M}$ , where  $M$  is equal to the total number of mutations in the dataset (Schürmann & Grassberger, 1996). Our results are robust with respect to the choice of  $\varepsilon$  (see below). The sum is taken over all  $N = 46$  genotypes and  $C_M = \frac{1}{N(1+\varepsilon)}$  is the normalization constant. The probability of a mutation *not* occurring in gene  $i$  is then  $P(\sigma_i = 0) = 1 - P(\sigma_i = 1)$ . We also estimate the joint probability distribution  $P(\sigma_i, \sigma_j)$  for each gene pair  $(i, j)$  as follows.



$$(4) \quad P(\sigma_i = 1, \sigma_j = 1) = C_J \sum_{g=1}^N \tilde{M}_{gi} \tilde{M}_{gj}$$

$$(5) \quad P(\sigma_i = 0, \sigma_j = 1) = C_J \sum_{g=1}^N (1 + \varepsilon - \tilde{M}_{gi}) \tilde{M}_{gj}$$

$$(6) \quad P(\sigma_i = 1, \sigma_j = 0) = C_J \sum_{g=1}^N \tilde{M}_{gi} (1 + \varepsilon - \tilde{M}_{gj})$$

$$(7) \quad P(\sigma_i = 0, \sigma_j = 0) = C_J \sum_{g=1}^N (1 + \varepsilon - \tilde{M}_{gi})(1 + \varepsilon - \tilde{M}_{gj})$$

where  $C_J = \frac{1}{N(1+\varepsilon)^2}$ . We use these estimates of joint probabilities to estimate the mutual information statistic  $MI_{ij}$  between random variables  $\sigma_i$  and  $\sigma_j$  as

$$(8) \quad MI_{ij} = \sum_{x,y \in \{0,1\}} P(\sigma_i = x, \sigma_j = y) \log_2 \frac{P(\sigma_i = x, \sigma_j = y)}{P(\sigma_i = x)P(\sigma_j = y)}$$

The aggregate mutual information statistic  $MI_{\text{tot}}$  for the full dataset is then calculated as

$$(9) \quad MI_{\text{tot}} = \sum_{i=1}^{K-1} \sum_{j=i+1}^K MI_{ij}.$$

### Generating null datasets

To obtain the null distributions for the individual  $MI_{ij}$  statistics and for the aggregate  $MI_{\text{tot}}$  statistic we generated “null” datasets that are structurally identical to our real data set, but in which the mutations are distributed randomly and independently across genotypes with the same marginal probabilities as in the real data. Specifically, in each “null” dataset, we generated  $N = 46$  genotypes by randomly and independently drawing each value  $M_{gi}, g = 1, \dots, N, i = 1, \dots, K$  from the Bernoulli distribution with

estimated marginal success probability  $P(\sigma_i = 1)$  for each gene  $i$ . This method preserves the average numbers of mutations per gene and per clone.

To obtain the null distributions for each  $MI_{ij}$  and  $MI_{tot}$ , we generated 100,000 “null” data sets, and calculated all  $MI_{ij}$  and  $MI_{tot}$  statistics for each “null” dataset as described above. We then estimated the  $p$ -value for all  $MI_{ij}$  and  $MI_{tot}$  and obtained nominally significant pairs of genes at different significance thresholds. Since the  $MI_{ij}$  statistics are not independent, we estimated the false discovery rate and the  $p$ -values for the observed number of nominally significant pairs from our “null” data sets (Kryazhimskiy, Sergey et al., 2011).

### **Strain construction**

Evolved alleles of the most significant gene pair, *PHO84* and *TRK1*, were reconstructed into the ancestral background using CRISPR-Cas9 allele swaps. We first constructed plasmids starting from pML104 (Addgene 67638), which constitutively expresses Cas9 and a guide RNA (gRNA). We designed gRNAs to target one site in *PHO84* (5' CCCGTAGAAAGCAACATCTAA 3') and two sites in *TRK1* (5' TTTTGGGTTCAAATCATCGAA 3' and 5' GGAGAACAACCTCTACTCGAC 3'). Plasmids were transformed into our ancestral background (yGIL1298: *MAT $\alpha$* , *ade2-1*, *CAN1*, *his3-11*, *leu2-3, 112*, *trp1-1*, *URA3*, *bar1 $\Delta$ ::ADE2*, *hmla $\Delta$ ::LEU2*, *GPA1::KanMX*, *ura3 $\Delta$ ::PFUS1-yEVENUS*) along with a 500 bp linear repair template (gBlock, IDT) encoding the appropriate evolved allele (*pho84*-A1071C, *trk1*-A733G, and *trk1*-C1353G) as well as a synonymous PAM site change. Transformants were genotyped to confirm successful integration of each mutant allele. The *pho84*-A1071C mutant strain was backcrossed to yGIL432 (*MAT $\alpha$* , *GPA1::NatMX*, otherwise isogenic to yGIL1298)

to move the *pho84*-A1071C allele to the *MATa* background. This *pho84*-A1071C *MATa* strain was crossed to yGIL1298 to generate heterozygous *pho84*-A1071C mutants, and to each of two *trk1* mutants to generate heterozygous double mutants. Heterozygous single *trk1* mutants were created by crossing correct transformants to yGIL432. All *MATa/α* diploids were then converted to *MATa/a* to correspond with the autodiploid background in which the mutations arose by transforming diploids with pGIL088, which contains a galactose-inducible *HO* homing endonuclease, as reported in (Fisher et al., 2018).

### **Fitness assay and interaction analysis**

Fitness assays were performed as described previously (Fisher et al., 2018). Briefly, mutant cultures were mixed 1:1 with an autodiploid version of the ancestral strain (yGIL1064) labeled with ymCitrine at *URA3*. Cultures were propagated in a 96-well plate in an identical fashion to the evolution experiment for 50 generations. At 10-generation intervals saturated cultures were sampled for flow cytometry. Analysis of flow cytometry data was performed with FlowJo 10.3. Selective coefficient was calculated as the slope of the best-fit line of the natural log of the ratio between query and reference strains against time.

Selection coefficients were measured for two technical replicates each of four biological replicates of *pho84*-A1071C and eight biological replicates of the remaining four query genotypes (*trk1*-A733G, *trk1*-C1353G, *pho84*-A1071C/*trk1*-A733G, and *pho84*-A1071C/*trk1*-A1353G). One reconstructed clone had an abnormally high fitness, likely due to secondary mutations introduced during transformation, and was removed from the analysis. There was no significant difference in fitness between two *trk1* alleles ( $t(28)=0.95$ ,  $p=0.35$ ) or two double mutants ( $t(28)=-1.087$ ,  $p=0.29$ ), so data for these

genotypes were pooled. The expected additive fitness distribution of the double mutant was calculated by adding the mean selection coefficients and propagating the standard deviation of *trk1* and *pho84* single mutants. A one-tailed two-sample t-test was used to test for deviation from additive expectation.

### **Network and clustering analysis**

Hierarchical clustering and heatmap generation were done using the pheatmap R package (Kolde., 2015). Mutual information matrices were clustered by rows and columns using a Euclidean distance matrix. Sub-clusters shown were identified by trimming row and column dendrograms to 5 groups and identifying the 4 sub-clusters containing less than 20 genes. The significant pair network was generated via using the R igraph package (Csardi & Nepusz, 2006).

## **Results**

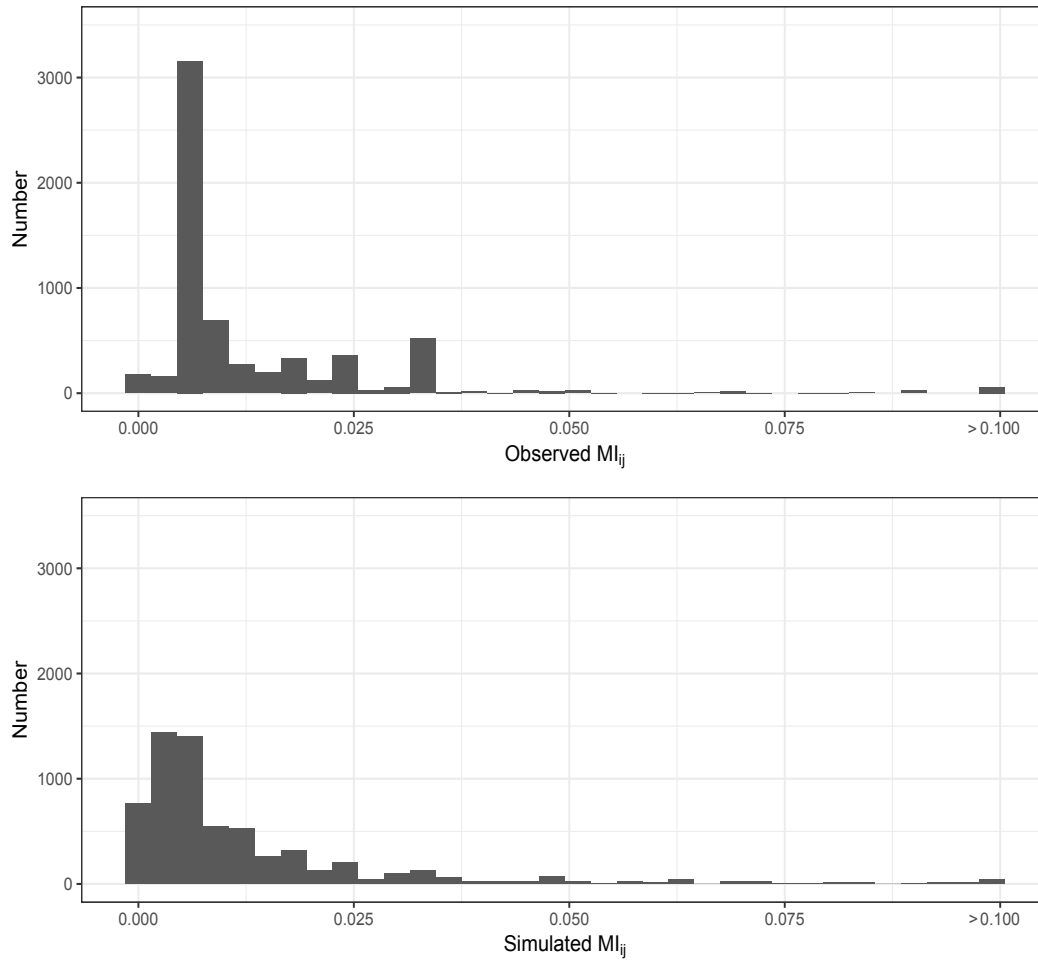
### **Identifying putative genetic interactions**

We set out to look for genetic interactions between beneficial mutations that arose in a previously published yeast evolution experiment for which whole-genome sequencing data are publicly available (Fisher et al., 2018). In this experiment, 46 replicate autodiploid yeast populations evolved in the same laboratory environment for 4,000 generations (Fisher et al., 2018). Using a custom bioinformatics pipeline (Methods), we identified 3,835 unique new mutations that arose during evolution. We found 113 “multi-hit” genes, i.e., genes in which a non-synonymous or a nonsense mutation was discovered in at least three independent populations. Since we expect to find only 13.5 of such genes by chance if mutations were distributed randomly across the genome, multi-hit genes must be highly enriched for targets of selection.

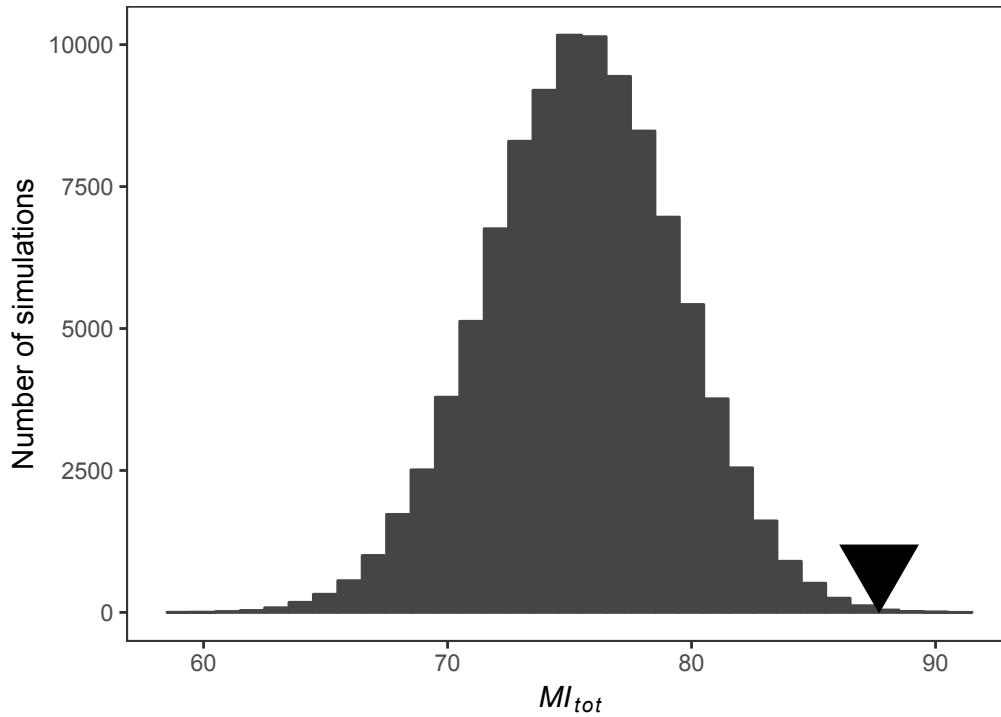
We asked whether any pairs of multi-hit genes occurred in our data more or less often than expected by chance. Such over- or underrepresentation would indicate parallel evolution driven by genetic interactions. We calculated the aggregate mutual information statistic,  $MI_{\text{tot}}$  (Methods), which serves as an overall measure of mutational non-independence in our dataset, and found that  $MI_{\text{tot}} = 87.7$  bits. We compared this value to the null distribution generated by randomly and independently distributing mutations among evolved genotypes  $10^5$  times (see Methods) and found that the observed value was significantly larger than expected by chance ( $p < 10^{-3}$ ; **Fig. 4-1, 4-2**). On average, the knowledge that a mutation in one gene is present in a given genotype provides a very small amount ( $87.7/6,328 = 0.014$  bits) of information about the presence of a mutation in any other specific gene. Nevertheless, the fact that mutated genes are distributed non-uniformly across evolved genotypes indicates that genetic networks subtly but significantly affected the mutational trajectories in our evolving populations.

Our estimates of mutual information depend on the value of the pseudocount parameter  $\epsilon$  (see Methods). We re-ran our analysis (albeit with 10 simulations instead of  $10^5$ ) at varying values of  $\epsilon$  between 0.0002 ( $\epsilon = \frac{0.1}{M}$ ) and 0.004 ( $\epsilon = \frac{2}{M}$ ) and found that our main result is robust with respect to the choice of  $\epsilon$  (**Fig. 4-3**).

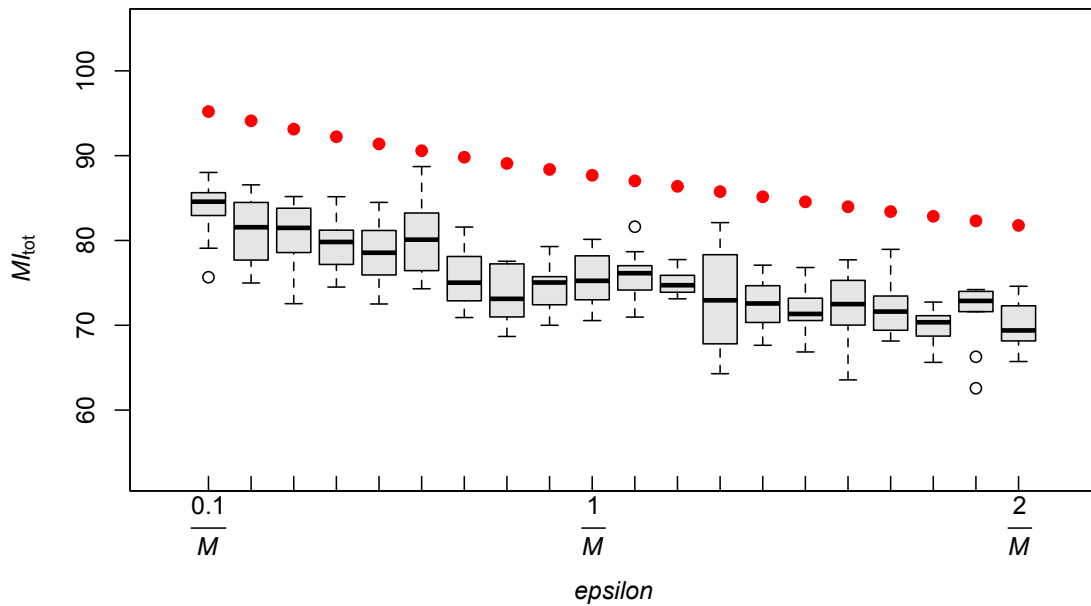
Next, we compared the mutual information statistic  $MI_{ij}$  for each gene pair ( $i,j$ ) in the dataset to its respective null distribution (see Methods, **Fig. 4-4**). We identified a significant genetic interaction between two genes if the  $p$ -value for their  $MI_{ij}$  was less than 0.003. At this cutoff, we expect to observe 18.8 interacting gene pairs under our null



**Figure 4-1.** Observed  $MI_{ij}$  values (top) visualized with a histogram and compared to a single randomly chosen simulation (bottom). Simulation histograms exhibit shifted distributions and reduced kurtosis.

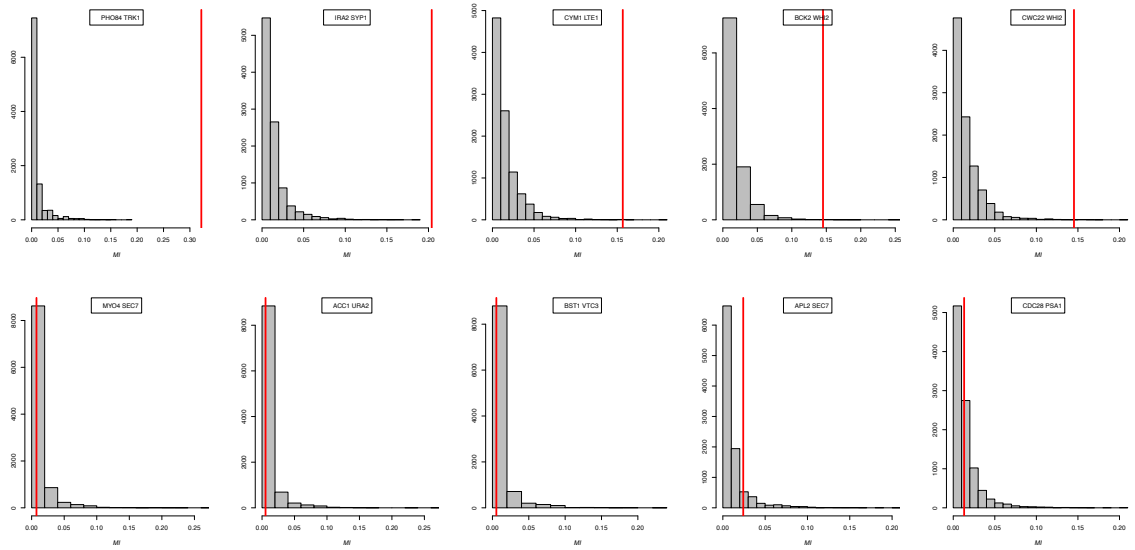


**Figure 4-2.** Histogram showing the null distribution of the aggregated  $MI_{tot}$  statistic based on 100,000 simulations (see Methods). The  $MI_{tot}$  observed in the real data is indicated by the black triangle. Observed genotypes contain significantly more information than expected by chance ( $p=0.00054$ ), indicating that interactions between mutations affect which *de novo* mutations accumulate in genotypes.



**Figure 4-3.** Our mathematical approach to calculating  $MI_{ij}$  in simulations necessitates a way to circumvent simulations in which gene  $i$ , gene  $j$ , or the double mutant ( $ij$ ) does not appear. We solved this by regularizing counts of  $M_i$  with a pseudocount parameter ( $\epsilon$ ). We measured observed and simulated  $MI_{tot}$  with  $\epsilon$  set to a range of values. Observed values are filled red dots and box and each box and whisker plot represents 10 simulations. The statistical difference between observed and simulated data is robust across this range of  $\epsilon$ .





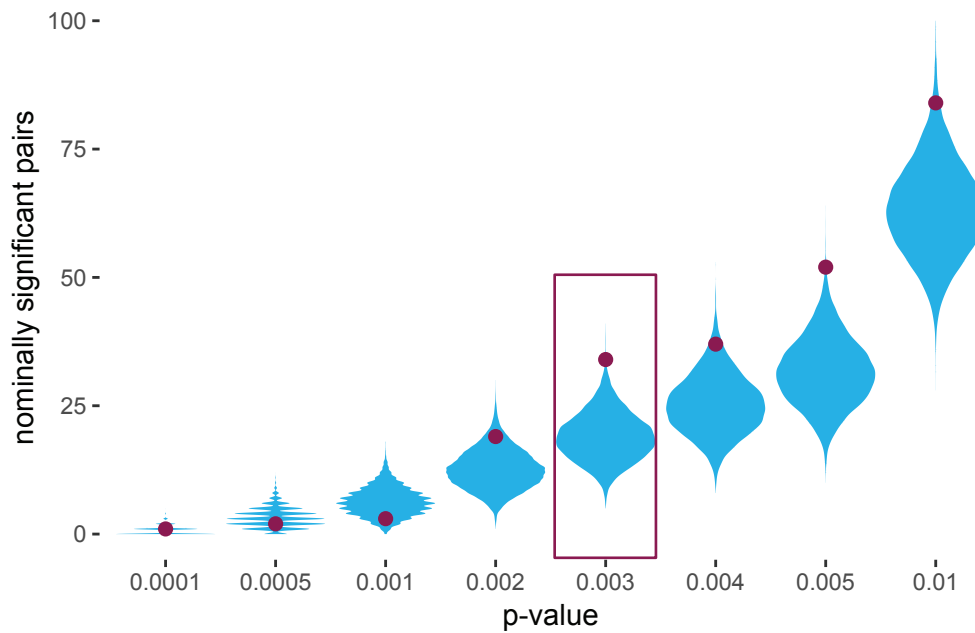
**Figure 4-4.** Examples of individual gene-pair  $MI_{ij}$  compared to respective simulated null distributions. Gene pairs in the top row are the five most highly significant interactions. Gene pairs in the bottom row are five randomly chosen non-significant pairs. Observed  $MI$  statistics are indicated with a red line and distributions based on a subsample of 10,000 simulations are plotted in grey.

model, but in fact we observe 33 (false discovery rate of 0.57) and this excess is highly significant ( $p < 0.005$ , **Fig. 4-5**). Thirty-three significant gene pairs are comprised of 42 unique genes (**Table 4-1**).

Interactions between functional variants might be expected to exhibit allele-specificity. We examined the identity of independently derived mutations in the top five most significant putative interactions. **Table 4-2** shows every incidence of an evolved mutation in the nine genes that participate in the top five most significant pairs identified above. Three of the nine genes in the top five pairs showed evidence of repeated loss of function as indicated by PROVEAN score (*IRA2*, *LTE1*, *WHI2*) (Choi & Chan, 2015). Mutations in the remaining six genes show a mix of predicted effects. We examined the positions of mutations within each gene to look for patterns of site-specific variation. We found that the distribution of mutations across coding sequences was consistent with uniform null hypothesis.

### **Experimental verification of genetic interaction between mutations in *PHO84* and *TRK1***

Despite the high false discovery rate, our epistasis analysis suggests that mutations in the top significant pair of genes, *PHO84* and *TRK1* (nominal  $p < 10^{-5}$ ) exhibit a true genetic interaction. Mutations in these two genes co-occurred in the same genotype in our data three times and never exhibited a higher value of mutual information in any of the 100,000 simulations. When examining the complete dataset, including all clones descending from each population, we found that a mutation in the *PHO84* gene precedes a mutation in *TRK1* in at least one population and that all populations with a non-synonymous mutation in *PHO84* allele acquire a *TRK1* mutation (**Fig 4-6**).



**Figure 4-5.** False discovery rate was estimated by sampling nominally significant pairs in a subset of 10,000 simulated datasets (blue violin plots). Observed significant pairs are plotted on top null distributions (purple). At  $p < 0.003$ , there are more significant pairs than expected by chance ( $p < 0.005$ ). An FDR of 0.57 was estimated by the dividing the mean number of nominally significant pairs by the number of observed significant pairs.

**Table 4-1.** Thirty-three nominally significant gene pairs.

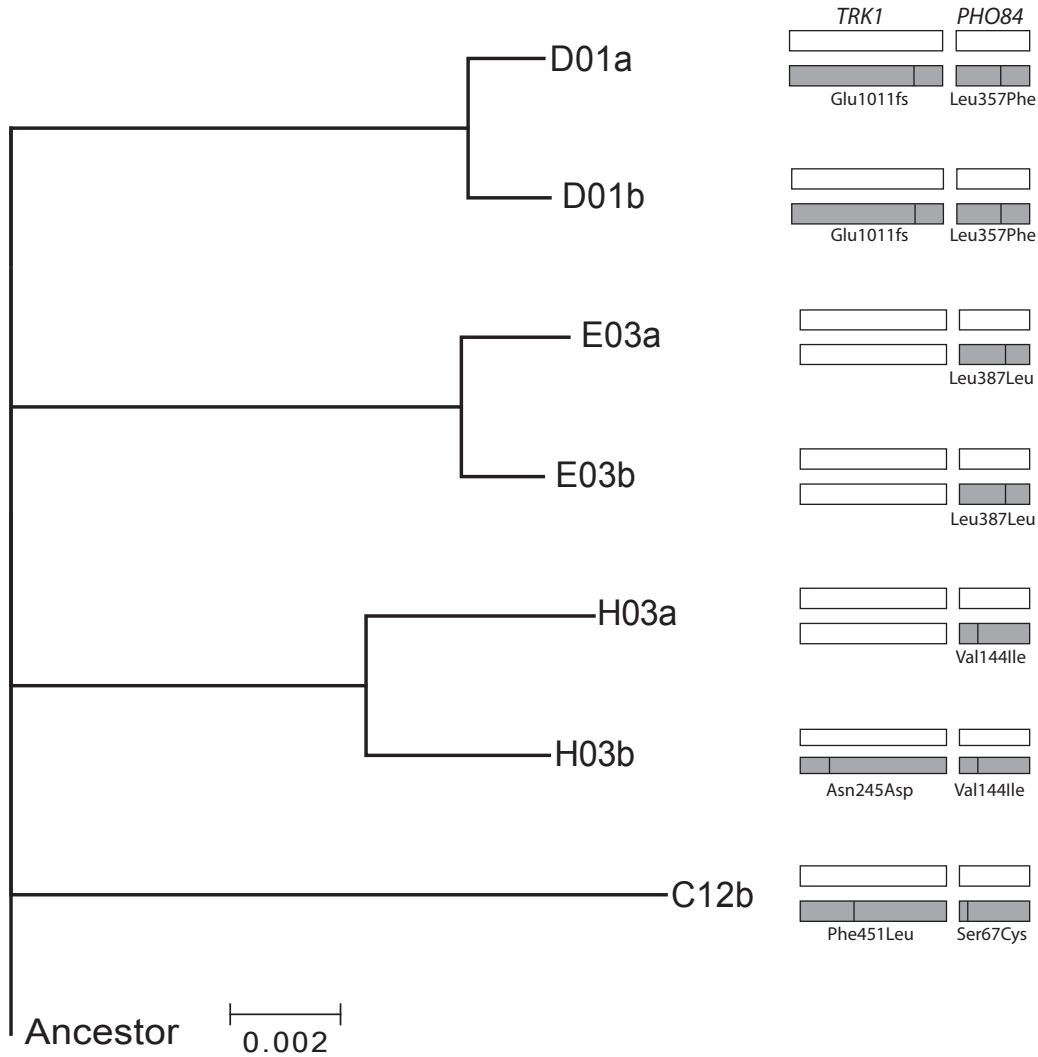
<b>Gene1</b>	<b>Gene2</b>	<b>Observed <i>I</i></b>	<b>Mean Simulated <i>I</i></b>	<b><i>p</i> Value</b>
<i>PHO84</i>	<i>TRK1</i>	0.321636423	0.009910122	<0.00001
<i>IRA2</i>	<i>SYPI</i>	0.20388364	0.013394146	0.000060
<i>CYM1</i>	<i>LTE1</i>	0.156670832	0.015274865	0.000860
<i>BCK2</i>	<i>WHI2</i>	0.145261454	0.015613055	0.000920
<i>CWC22</i>	<i>WHI2</i>	0.145261454	0.015672584	0.000980
<i>NOC2</i>	<i>YBT1</i>	0.133466802	0.009807874	0.001520
<i>DTR1</i>	<i>TOM1</i>	0.133466802	0.009775809	0.001590
<i>KEL1</i>	<i>YBT1</i>	0.133466802	0.009800946	0.001750
<i>GRR1</i>	<i>SIN4</i>	0.133466802	0.009904341	0.001760
<i>HIS4</i>	<i>TRK1</i>	0.133466802	0.009849421	0.001760
<i>RSE1</i>	<i>YLR089C</i>	0.133466802	0.009841149	0.001780
<i>MAK21</i>	<i>SIN4</i>	0.133466802	0.00997965	0.001820
<i>MNN9</i>	<i>SRS2</i>	0.133466802	0.00996807	0.001820
<i>NOC2</i>	<i>URA2</i>	0.133466802	0.009847824	0.001840
<i>SIN3</i>	<i>tL(GAG)G</i>	0.133466802	0.00987234	0.001880
<i>PHO84</i>	<i>SIN3</i>	0.133466802	0.009921999	0.001910
<i>MNN9</i>	<i>SOG2</i>	0.133466802	0.00984068	0.001920
<i>PMD1</i>	<i>TPS2</i>	0.133466802	0.009930568	0.001950
<i>BST1</i>	<i>TRAI</i>	0.133466802	0.009838169	0.001960
<i>SIN3</i>	<i>TRK1</i>	0.133466802	0.00983419	0.001960
<i>AVO1</i>	<i>PHO4</i>	0.133466802	0.009952229	0.001970
<i>SMI1</i>	<i>TRAI</i>	0.133466802	0.009881812	0.001990
<i>JSN1</i>	<i>YMR144W</i>	0.133466802	0.009930848	0.002000
<i>NOC2</i>	<i>PCAI</i>	0.133466802	0.009875012	0.002040
<i>PCAI</i>	<i>YBT1</i>	0.133466802	0.009869169	0.002050
<i>HIS4</i>	<i>PHO84</i>	0.133466802	0.009950975	0.002100
<i>SOG2</i>	<i>SRS2</i>	0.133466802	0.009866437	0.002100
<i>ACC1</i>	<i>SYPI</i>	0.133466802	0.009878601	0.002120
<i>AVO1</i>	<i>RTF1</i>	0.133466802	0.009929195	0.002130
<i>KRR1</i>	<i>MNN9</i>	0.133466802	0.009938925	0.002130
<i>PCAI</i>	<i>YMR144W</i>	0.133466802	0.009969454	0.002190
<i>SRS2</i>	<i>YDR119W</i>	0.133466802	0.009908157	0.002240
<i>CSF1</i>	<i>IRA1</i>	0.126589134	0.015444545	0.002670

**Table 4-2.** Evolved alleles of genes in top 5 putative interactions.

Population/Clone	PHO84	TRK1	
pop C12	Ser67Cys (N)	Phe451Leu (N)	
pop D01	Leu357Phe (D)	Glu1011fs (D)	
pop E03	Leu387Leu (N)	wt	
clone H03a	Val144Ile (N)	wt	
clone H03b	Val144Ile (N)	Asn245Asp (N)	
	IRA2	SYP1	
clone A03b	Phe2557fs (D)	Gln180Glu (N)	
pop B01	Glu2774* (D)	wt	
pop C11	Ile1463fs (D)	Gly300fs (D)	
pop D01	wt	Ile753Ile (N)	
clone D10a	Glu2440* (D)	wt	
clone D12a	Ser1292Phe (D)	wt	
clone D12b	Ser1292Phe (D)	Val126Leu (N)	
pop H03	Lys2651Asn (N)	wt	
	CYM1	LTE1	
pop B03	wt	Ser185* (D)	
pop C02	wt	Lys1138* (D)	
pop C03	wt	Ala1368fs, Ala748Ser (D)	
pop C11	His700Asp (N)	Ala748Val (N)	
pop D12	wt	Glu653* (D)	
pop F02	Asp145Val (D)	Glu865* (D)	
pop F03	wt	Met1062Ile, Gln916Lys (D)	
clone F10a	wt	Trp380* (D)	
clone F10b	Leu274Ser (N)	Trp380* (D)	
clone G02a	wt	Trp380* (D)	
clone G12b	wt	Trp1403* (D)	
	BCK2	WHI2 \$	CWC22
pop C12	Asp452His(N)	Thr283fs†(D)	wt
pop D03	wt	Ala338* (D)	wt
pop E03	Asn217fs (D)	Leu76fs† (D)	wt
pop E11	wt	Gln81* (D)	Val60Val (N)
clone E12a	wt	Gln181*† (D)	Wt
clone E12b	wt	Gln181*(D)	wt
clone F11a	wt	Gln29*† (D)	Leu526Phe (N)
clone F11b	wt	Gln29*† (D)	wt
pop G01	wt	Glu168Gly (D)	wt
clone G02a	Arg370*, Pro57Leu (D)	Ser72*† (D)	wt
clone G02b	Arg370* (D)	Ser72*(D)	wt
pop G11	wt	Asn275fs(D)	Cys245Gly (N)
clone H11a	wt	Leu76fs†, Ala310Pro (D)	Thr183Pro (N)
clone H11b	wt	Leu76fs, Ala310Pro (D)	Thr183Pro (N)

All mutations to genes in the top 5 significant pairs are listed along with their PROVEAN predicted effect (cutoff = -2.5) ; (N) neutral, (D) deleterious. \$WHI2 putatively interacts

with both CWC22 and BCK2. † Mutations are homozygous. Fixed alleles are identified by population and non-fixed alleles are identified by clone. All genotype data published previously in Fisher et al. (2018).



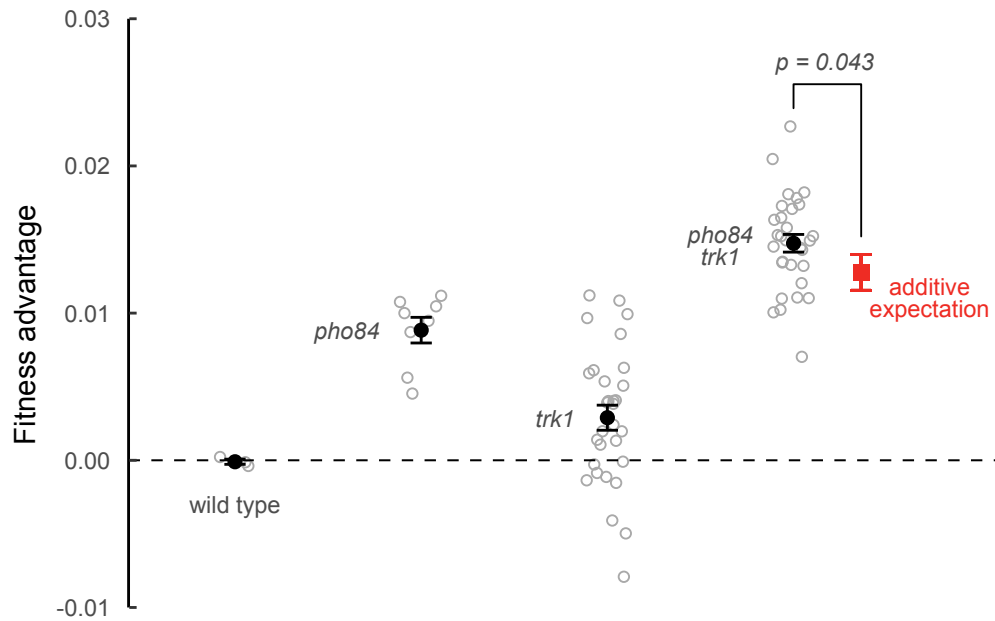
**Figure 4-6.** All autotetraploid isolates containing a *TRK1* or a *PHO84* mutation. Phylogeny is a ML tree based on 3,835 variable sites. Sequence data was obtained for 2 clones in all populations except C12. The only population in which a *pho84* mutation is not accompanied by a mutant *trk1*, E03, also contains the only synonymous *pho84* allele. The only population polymorphic for *TRK1*, H03, indicates that the *PHO84* mutation arose first.

To experimentally validate this positive genetic interaction, we reconstructed one allele of *pho84* and two alleles of *trk1* in the ancestral background both as single mutants and as double *pho84/trk1* mutants. All mutations were constructed as heterozygotes—the state in which they are maintained in the evolution experiment—and assayed for fitness. The mutant *pho84* and *trk1* alleles conferred small, but measurable fitness benefits ( $0.009 \pm 0.002$  s.d. for *pho84* and  $0.003 \pm 0.005$  s.d. for *trk1*). We found that the fitness of the *pho84/trk1* double mutant ( $0.015 \pm 0.003$  s.d.) was higher than the expectation based on the sum of fitnesses of single mutants ( $0.013 \pm 0.005$  s.d., **Fig. 4-7**), although the difference was only marginally significant ( $t(58)=1.74, p=0.043$ ).

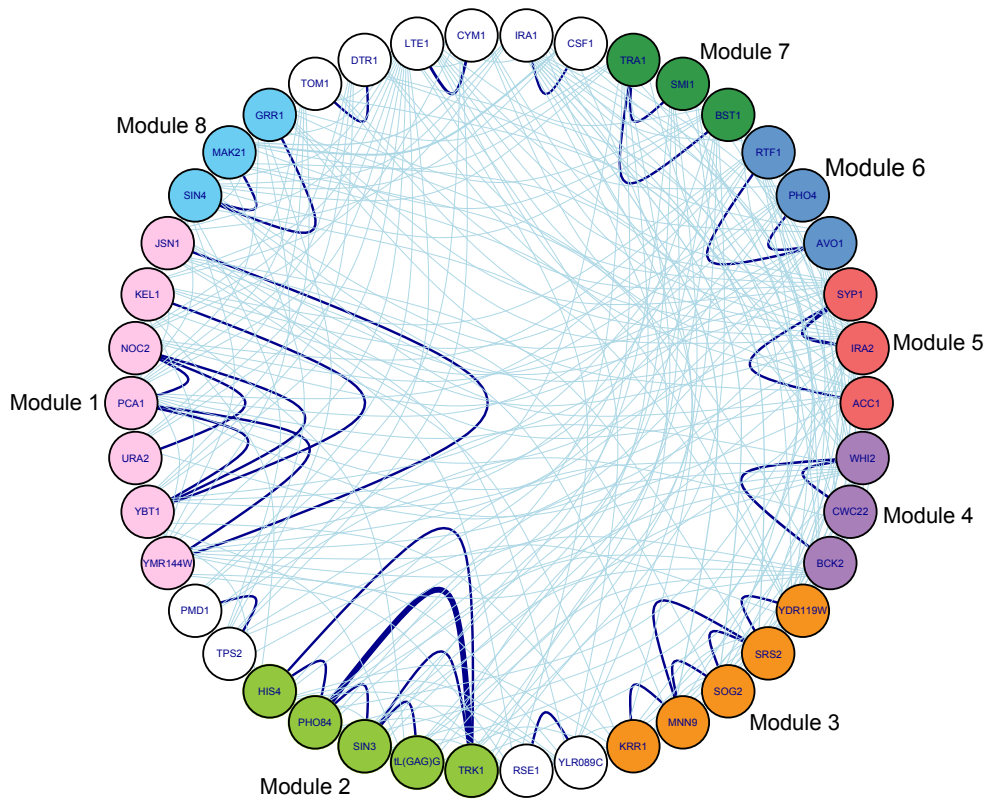
### ***Structure of genetic-interaction networks***

We found that the set of putatively interacting genes is highly interconnected. The most significant 33 gene pairs consist of 8 modules that contain at least three genes each and five isolated gene-gene interaction pairs (**Fig. 4-8**). The three largest modules encompass 42% of all candidate significant interactions. We performed hierarchical clustering on  $MI_{ij}$  and found that this matrix contains multiple small but tightly connected blocks (**Fig. 4-9**). On average, mutual information between any two genes within a block was 8 times higher than between a random pair of genes (0.098 bits vs 0.012 bits). Notably, these blocks largely overlapped with the modules observed among putatively interacting pairs. This suggests that genetic interactions, rather than being exclusively strong pairwise interactions, are often dispersed among small networks of interacting genes.

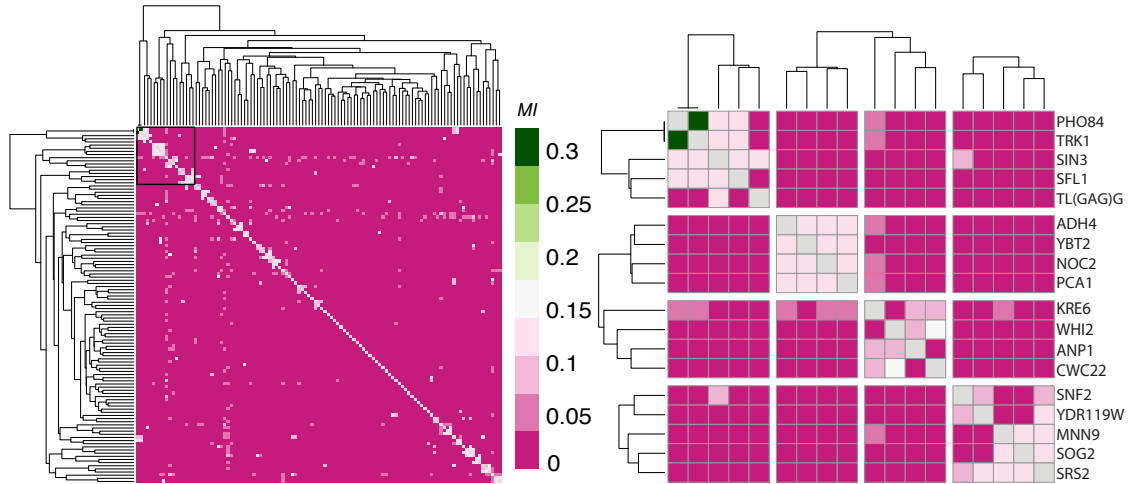




**Figure 4-7.** Fitness advantage of the single *TRK1* and *PHO84* mutations and of the double mutant. Fitness measurements for two *TRK1* alleles are combined. Replicate measurements are plotted as grey circles. Mean estimates are plotted as bold circles  $\pm$  standard error. The red square indicates the additive expectation for the double mutant. Additive fitness expectation is the sum of the mean fitness measurements for both single mutants. Standard error for the double mutant is propagated SE from single mutant replicates.



**Figure 4-8.** Network of all genes identified in significant gene pairs. Edges are scaled by  $MI_{ij}$  and connect all genes that co-occur in the same background at least once. Bolded lines represent significant pairwise  $MI_{ij}$ . Colors correspond to interconnected significant pairs. White circles indicate isolated gene pairs. Modules are labeled by size.



**Figure 4-9.** Hierarchical clustering of genes by pairwise mutual information captures the most significant pairs and networks among significant pairs. Sub-clusters shown were identified by trimming row and column dendrograms to 5 groups and identifying the 4 sub-clusters containing less than 20 genes.

## Discussion

Even the simplest free-living microorganisms encode thousands of genes organized into complex and interconnected networks that collectively determine the organism's fitness. These genetic interaction networks constrain evolution such that populations evolving in identical conditions often find similar genetic solutions, both in nature and in the laboratory (e.g. (Kvitek & Sherlock, 2013; Protas et al., 2006; Tenaillon et al., 2012). Here we developed a method, based on mutual information, that exploits genetic parallelism observed in microbial evolution experiments to infer genetic interactions between loci that acquired mutations in independent populations. With this method, we found that genetic interactions had an overall significant effect on mutational trajectories of evolved populations. We also identified 33 gene pairs (at FDR of 0.57) that exhibit the strongest genetic interactions in our data set. We provide experimental support for one of these interactions, between genes *PHO84* and *TRK1*.

Our method for detecting genetic interactions complements existing approaches. Most of our understanding of genetic interactions comes from the systematic analysis of double deletion/knockdown mutations (Babu et al., 2014; Baryshnikova et al., 2010; Breslow et al., 2008; Costanzo et al., 2010; Costanzo et al., 2016; Jasnos & Korona, 2007; Szappanos et al., 2011; Tong et al., 2004; Van Opijnen, Bodi, & Camilli, 2009). By design, these approaches query only loss-of-function mutations, which represent less than 5% of natural variation in both yeast and humans (Bergström et al., 2014; Saleheen et al., 2017). In contrast, our approach can detect pairwise epistasis between all classes of beneficial variants, including gain-of-function mutations, mutations in essential genes, and regulatory mutations, that would be missed in gene-deletion studies. Indeed, out of the 33 most significant gene pairs only one genetic interaction (between *SIN3* and *TRK1*)

was known previously (Costanzo et al., 2016). The most significant interaction discovered here is between genes *TRK1* and *PHO84*. *TRK1* encodes a high-affinity potassium transporter and *PHO84* encodes a high-affinity phosphate transporter. The biological cause of their interaction is unclear, although there is evidence of crosstalk between potassium and phosphate import (Canadell, González, Casado, & Ariño, 2015; Rosenfeld et al., 2010).

While gene-deletion studies are particularly good at detecting strong negative pairwise interactions between deleterious mutations, such as synthetic lethality (reviewed in (Baryshnikova et al., 2010), our method identifies primarily positive interactions between pairs of selectively accessible mutations. In theory, our approach could also capture negative interactions, but this would require observing an absence of certain mutational combinations more often than expected by chance. Such mutational incompatibilities have been observed in evolution experiments (Kvitek & Sherlock, 2013; Tenaillon et al., 2012); for example mutations in a *HXT6/7* hexose transporter and its negative regulator, *MTH1*, in glucose-limited yeast chemostat populations are incompatible (Kvitek & Sherlock, 2011). The absence of negative interactions in our list of significant pairs suggests that we are underpowered to detect them.

Several recent experimental evolution studies have found that adaptive mutations often exhibit a global (i.e. not specific to a particular gene pair) type of negative epistasis, which is referred to as “diminishing returns epistasis” (Chou, Chiu, Delaney, Segre, & Marx, 2011; Echenique, Kryazhimskiy, Ba, & Desai, 2019; Khan, Dinh, Schneider, Lenski, & Cooper, 2011; Kryazhimskiy, S., Rice, Jerison, & Desai, 2014). For example, we previously demonstrated that beneficial alleles in *gas1* and *ste12*, both ~3% fitness

effect mutations, yield only a net ~5% benefit when combined (Lang et al., 2013). If diminishing returns epistasis is indeed widespread, then pairwise interactions between specific mutations should be detected as deviations of the double-mutant fitness from the appropriate diminishing returns null model rather than from a naïve additive model. Then, observing a double-mutant with higher-than-additive fitness, such as the *TRK1/PHO84* double mutant (**Fig. 4-7**), would be even more surprising compared to the diminishing returns null than to the additive null, and would provide even stronger evidence for a gene-specific positive genetic interaction.

Our approach has several important limitations. It suffers from a high rate of false discoveries (about 60%), at least for the dataset that we have analyzed here. There are at least two reasons for such high FDR. First, we looked for signatures of epistasis among pairs of genes in which we observed three or more independent mutations. We assumed that all observed mutations in these “multi-hit” genes are beneficial. However, this may not be the case. We estimate around 12% of the genes included in this analysis to have been mutated three or more times simply by chance. These mutations are distributed uniformly among genotypes and therefore decrease the signal to noise ratio in our data. One way to decrease FDR is to consider genes with an even higher degree of parallelism. Of course, this would come at a cost of potentially missing interesting genetic interactions among less frequently mutated genes.

Second, high FDR may in fact reflect a real biological phenomenon. Gene deletion studies have shown that strong pairwise epistasis is relatively rare, around 4% if both positive and negative interactions are counted (Costanzo et al., 2016). Thus, strong pairwise genetic interactions among beneficial mutation might also be rare. Weak

epistasis might be more common, but it is also harder to detect. The highly significant value of the aggregate  $MI_{\text{tot}}$  statistic in our study suggest that genetic interactions jointly have affected the outcome of the evolutionary process at the genetic level. At the same time, the difficulty of reliably identifying individual interacting gene pairs suggests that genetic interactions, rather than being strong and concentrated in a small number of gene pairs, are weak and relatively dispersed. The power of our approach to detect weaker genetic interactions could be improved with more replicate populations. In our null model, co-occurrence of two mutations in the same genotype happens with probability on the order  $N^{-1}$ , where  $N$  is the number of independently evolved genotypes. For example, the  $p$ -value for two genes with three mutations each where all mutations co-occur in the same three genotypes scales as  $N^{-3}$ .

As mentioned above, our method is designed to detect pairwise genetic interactions. However, we observe that putative genetic interactions that we identify are clustered in groups that contain two to seven genes. It is tempting to conclude that such clustering is caused by real biological modules corresponding to physiologically distinct routes of adaptation. However, some degree of clustering is expected even if all of epistasis were pairwise and uniformly distributed among genes. The amount of such spurious clustering would depend on the strength and prevalence of epistasis and is hard to estimate. Increasing the number of replicate populations and reducing the duration of evolution experiments is likely to alleviate this problem.

Our approach does not eliminate the need for experimental validation of putative genetic interactions. However, current molecular techniques make genetic reconstructions feasible only for a relatively small number of mutations. Thus, our approach could serve

as an initial filter for narrowing down the set of potentially interesting pairs of mutations for further experimental validation and investigation.

Our results demonstrate the feasibility of using experimental evolution and genetic parallelism to identify biologically interesting genetic interactions that might otherwise be difficult to uncover. In combination with other approaches, it will facilitate characterization of epistasis and, more broadly, help us understand the factors driving patterns of parallelism, diversification, and genomic constraint in evolution.



# Appendix A

## Statistical method for detection of genes harboring beneficial mutations

### **Summary**

Many findings from chapters one and two of this dissertation rely on the ability to detect genes that are parallel targets of selection in yeast evolution experiments. This is accomplished by identifying those genes in the yeast genome that receive more mutations during an evolution experiment than expected by chance. This is not a statistically trivial problem given that most genome evolution is neutral (Buskirk et. al 2017). To solve this problem, we developed a model to determine the probability that observed number of mutations in each specific gene occurred by chance. This model was used in Fisher & Buskirk et al. (2018) to identify genes harboring beneficial mutations. I then used this model in unpublished work to compare common targets of selection between three different evolution experiments, as well as look for beneficial intergenic mutations. In this appendix I will introduce the model and describe the latter two analyses.

### **Materials and Methods**

#### ***Genomic datasets***

We identify beneficial variants and pairwise interactions in three previously published datasets of experimentally evolved *S. cerevisiae*: 1,000-generation evolved haploid (Lang et al., 2013), 2,000-generation evolved diploids (Marad et al., 2018), and 4,000-generation evolved autodiploid (Fisher et al., 2018). These three yeast lines were all evolved with different genomic backgrounds; haploids of mating-type **a** (*MATa*), autodiploids homozygous for mating type (*MAT $\alpha$ / $\alpha$*  & *MATa/a*), and true diploids heterozygous for mating type (*MATa/ $\alpha$* ), thus the three sets of populations all exhibit different combinations of mating-type and ploidy. Both mating-type and ploidy contribute to differences in gene regulation, therefore these three experimental lines have appreciable differences in genomic background.

### ***Sequencing data re-analysis***

All sequencing data used here were previously reported (Lang et al., 2013; Marad et al., 2018; Fisher et al., 2018). However, because datasets were analyzed years apart, all data was reanalyzed with a common pipeline to ensure comparability. All raw data files were demultiplexed using a custom python script (barcodesplitter.py) from L. Parsons (Princeton University). Adapter sequences were trimmed using fastx\_clipper (FASTX Toolkit). Reads were then aligned to a customized W303 genome using BWA v0.7.12 (Li et al., 2009). VCFtools was used to filter variants common to all samples and mating-type specific polymorphisms. Remaining polymorphisms were then annotated using a strain-background customized annotation file (Matheson et al. 2017). Intergenic variants were further annotated based on flanking genes using in-house scripts. The data generated in Lang et al. (2013) are from whole population sequencing while the other two datasets are from clone sequences. To generate comparable genotype data for common

target detection, we identified dominant lineages in each population. Each lineage consists of variants in the same clonal background, and is therefore equivalent to clone genotypes.

### ***Identification of parallel targets of selection***

We identify putative genetic targets of selection by identifying genes in which we observe more mutations than we expect by chance. Assuming that mutation events occur independently and at a constant rate, the number of mutations in a defined stretch of the genome will follow the Poisson distribution. We calculate the expected number of mutations for each gene or intergenic region,  $\sigma$ , using the Poisson distribution weighted for the length,  $L$ , of each region:

$$(1) \quad \lambda_{\sigma} = \left( \frac{L_{\sigma}}{\sum_{\sigma}^N L_{\sigma}} \right) M$$

where  $M$  is the total number of mutations in the dataset. The probability of observing  $k$  mutations in gene  $\sigma$  is therefore

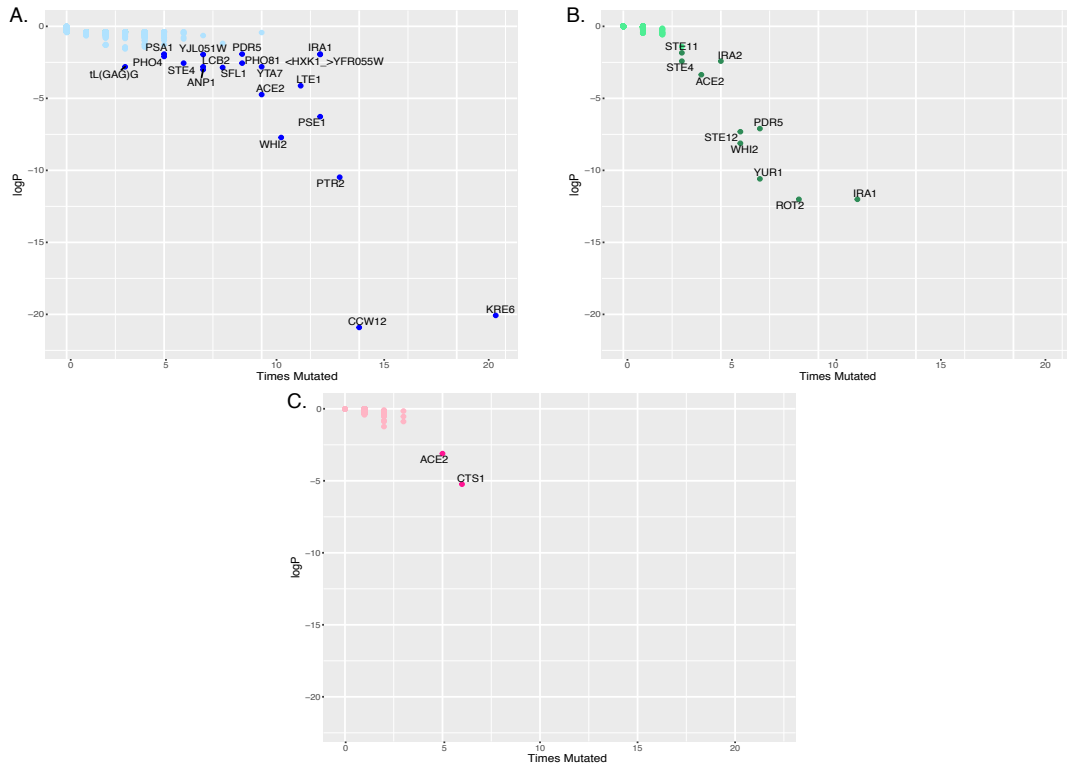
$$(2) \quad P\{obs = k\} = \frac{\lambda_{\sigma}^k}{k!} e^{-\lambda_{\sigma}}$$

We divided the entire yeast genome into 6,264 genic and 6,243 intergenic windows by length in base pairs. We use expression (2) to calculate the  $p$ -value for the observed number of mutations in each genomic window. We then applied the Benjamin-Hochberg post hoc adjustment to correct for multiple hypothesis testing.

## **Results**

### ***Identifying putative genic and intergenic targets of selection***

Each set of mutations was independently surveyed for overrepresented gene and intergenic targets (**Fig. A-1**). Thirty targets comprised of twenty-five unique genes or



**Figure A-1.** Loci with parallel mutations in three datasets. Twenty-four unique genes and one intergenic region were identified as overrepresented in whole-genome sequence data with a significance threshold of  $p < 0.02$ . Parallelism is seen both within and between evolution experiments with different genetic backgrounds. (A) Autodiploid populations; blue, (B) haploid populations; green, (C) diploid populations; pink.

intergenic regions were identified as overrepresented in whole-genome sequence data with a significance threshold of  $p < 0.02$ . Most genes identified here have been previously reported to contain beneficial variants. The inclusion of intergenic units in this analysis yielded one overrepresented non-coding region, the intergenic stretch between *HXK1* and *IRC7*. Although *de novo* adaptive non-coding variants have been described in bacterial systems (Solopova et al., 2012; Toprak et al., 2012), none have been identified in experimentally evolved yeast populations. *HXK1* and *IRC7* are divergently transcribed, thus variants in this region are upstream of both genes. It is unlikely, however, that these variants affect regulation since the 13 observed variants are distributed throughout this 10 kb region.

## **Discussion**

Most experimental evolution studies identify genes that receive mutations in more than one replicate population and interpret this as evidence of selection at those loci (Lang et al, 2013; Gorter et al., 2017; Behringer et al., 2018; Marad et al., 2018). Parallelism is an effective way to identify beneficial mutations, however, simply identifying genes that receive mutations in more than one independent line is a less than ideal means of identifying targets of selection, and is likely to generate many false positives due to widespread hitchhiking in clonally propagated experimental populations (Lang et al., 2013, Buskirk et al, 2017). We created a simple statistical method to calculate the probability that each gene in the genome received an observed number of mutations during a given experiment. This is a much more rigorous means of identifying loci harboring beneficial mutations.

A strength of this approach is the ability to compare commonly mutated genes across experiments using different genomic backgrounds. We reanalyzed three published datasets of yeast experimental evolution: haploids, diploids, and autodiploids, cells that have two copies of the genome but are physiologically haploid because they are homozygous at the mating-type locus. These three datasets have previously not been compared in a systematic way, nor have intergenic variants been included in overrepresentation analyses. Although the experimental populations here originate from the same laboratory strain, background dependence with regards to ploidy and the mating-type (*MAT*) locus can still clearly be seen. Some genes appear enriched in specific datasets, such as *ROT2* and *STE12* in haploids. Other genes, such as *ACE2*, are ubiquitously accessible targets for beneficial variants.

Although regulatory variation is expected to play a role in adaptation, particularly in diploid populations (Wray, 2007), no individual intergenic sequence has previously been found to significantly accumulate variants in yeast experimental evolution. Across the three datasets, we find only a single intergenic region, a 10 kb region upstream of both *HXK1* and *IRC7*, to be overrepresented. Unlike genes, where it is straightforward to define a functional unit, promoters are often small and poorly defined. It is likely; therefore, that examining overrepresentation in entire intergenic units could obscure adaptive regulatory variants due to the small functional target size. The 13 observed variants are distributed throughout this 10 kb region upstream of both *HXK1* and *IRC7*.

The weakness of this approach is that it is dependent on the total number of variants observed in an experiment. The total number of mutations identified in an experiment will depend on the number of replicate populations, the population size, the

duration of the experiment, and the strength of selection. This method is most effective for experimental conditions in which beneficial mutations saturate. This does not appear to happen in yeast evolution experiments and therefore the list of genic targets of selection identified by recurrence alone is likely incomplete.

***Code availability***

The code to run the statistics described above is archived in the Lang Lab github at [github.com/LangYeastEvoLab/Useful\\_Scripts](https://github.com/LangYeastEvoLab/Useful_Scripts).

# Appendix B

## Supporting material for chapter 2

### Supplemental Methods

#### Strain construction

To generate deletion strains, a linear hygromycin-resistance cassette was amplified with overhanging 40 bp of *STE4* homology on both ends and transformed into yGIL1113 (*MAT $\alpha$* , *ura3 $\Delta$* ::*P<sub>Fus1</sub>*-yEVENUS; Lang et al. 2011). The resulting deletion mutant was converted to *MAT $\mathbf{a}$*  by transforming with a gal-inducible *HO* (as described below) and used as the haploid deletion strain in competitions. The same deletion mutant was also crossed to yGIL121 (*STE4*, *ura3*) and yGIL128 (*ste4 $\Delta$* ::*KanMX*, *ura3*; reported as DBY15087 in Lang et al. 2009) to generate heterozygous and homozygous deletions. A cross of yGIL1113 to yGIL121 was used as a wild type control. Crosses of strains carrying null alleles were performed by first transforming with a *STE4*-expressing plasmid from the MoBY ORF plasmid collection to complement *ste4 $\Delta$* . All diploid genotypes were then converted from *MAT $\mathbf{a}$* / *$\alpha$*  to *MAT $\mathbf{a}$* / *$\mathbf{a}$*  via transforming with pGIL088, a plasmid encoding a gal-inducible *HO* and a *P<sub>STE2</sub>**HIS3* marker, and selecting his<sup>+</sup> transformants.

Three evolved *STE4* alleles were selected for reconstruction, 81 $\Delta$ T (Ser261fs), G943T (Glu315\*), and G935A (Arg312Gln). Alleles were reconstructed in yGIL432 (a yGIL121 derivative) using CRISPR-Cas9 alleles swaps. We constructed constitutive



Cas9-expressing plasmids starting from pML104 (Addgene 67638) expressing a *STE4*-specific guide RNAs (5' CTACCCCTACTTATATGGCA 3') and co-transformed the plasmid along with a 500 bp linear repair template (gBlock, IDT) encoding the one of three evolved alleles as well as a synonymous C954A PAM site substitution. A strain containing just the synonymous PAM site was also isolated and used as the wild-type control allele. To minimize variation due to transformation and Cas9 activity, one successful transformant per allele was backcrossed twice and the resulting diploid was sporulated and tetrad dissected. For each allele, spores were genotyped at *STE4* and intercrossed to generate heterozygous and homozygous mutants. Crosses of strains carrying evolved *ste4* alleles were performed by first transforming with a plasmid from the MoBY ORF plasmid collection to complement *STE4*. Mutants carrying an evolved *WHI2* C85T (Q29\*) allele were generated in identical fashion with two exceptions. The evolved substitution is within the *WHI2* gRNA used (5' ACAGTACGAAGGTAACGAGG 3'), and therefore no synonymous mutation was needed to eliminate Cas9 activity. A correct *whi2*Q29\* was backcrossed once and intercrossed to produce homozygotes and heterozygotes. All diploid genotypes were converted to *MATa/a* as described above.

We generated strains containing dominant drug cassettes tightly linked to the *WHI2* locus using a similar CRISPR-based approach. We inserted either HphMX or KanMX 220 bp downstream of *WHI2* or *whi2*C85T by transforming with the same gRNA (5' ATCCCCTTCTGCAAATAACG 3') and Cas9-expressing plasmid and co-transforming with linear drug cassettes flanked by 40 bp of homology to the targeted region. Successful transformants were then backcrossed to either a wild type background

or a *ste4G943T* (described above) mutant. Crosses were sporulated and spores were selected in which drug-marker tagged mutant and wild-type *WHI2* alleles are present on the same chromosome as both mutant and wild type *STE4*. Correct spores were crossed to generate three genotypes; *WHI2::HphMX-STE4/WHI2::KanMX-STE4*, *WHI2::Hph-STE4/whi2C85T::KanMX-STE4*, and *WHI2::Hph-STE4/whi2C85T::KanMX-ste4G943T*. All three genotypes were converted to *MAT<sub>a/a</sub>* as described above.

**Table B-1.** Strains generated for this study.

Ploidy	Assay	Genotype*†
<i>MATa</i>	Competitive fitness	<i>STE4, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa</i>	Competitive fitness	<i>ste4 Δ::KanMX, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>STE4, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>STE4/ ste4 Δ::KanMX, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4 Δ::KanMX, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4 Δ::KanMX, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4 Δ::KanMX, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa</i>	Competitive fitness	<i>ste4C958A<sup>‡</sup>, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4C958A<sup>‡</sup>/STE4, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/ GPA1::KanMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4C958A<sup>‡</sup>/ ste4C958A<sup>‡</sup>, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/ GPA1::KanMX</i>
<i>MATa</i>	Competitive fitness	<i>ste4T81Δ, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4T81Δ/STE4, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/GPA1::KanMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4T81Δ, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/GPA1::KanMX</i>
<i>MATa</i>	Competitive fitness	<i>ste4G943T, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4G943T /STE4, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/GPA1::KanMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4G943T, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/GPA1::KanMX</i>
<i>MATa</i>	Competitive fitness	<i>ste4G935A, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4G935A /STE4, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/GPA1::KanMX</i>
<i>MATa/a</i>	Competitive fitness	<i>ste4G935A, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus/ura3, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX/GPA1::KanMX</i>
<i>MATa/a</i>	Short-term-evolution	<i>STE4-WHI2::HphMX/ STE4-WHI2::KanMX,</i>

		<i>ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Short-term-evolution	<i>WHI2::HphMX-STE4/whi2C85T::KanMX-STE4, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>
<i>MATa/a</i>	Short-term-evolution	<i>WHI2::HphMX-STE4/whi2C85T::KanMX-ste4G943T, ade2-1, his3-11, leu2-3,112, trp1-1, ura3Δ::P<sub>Fus1</sub>-yEVENus, bar1Δ::ADE2, hmlaΔ::LEU2, GPA1::NatMX</i>

\*Two alleles provided only for heterozygous loci in diploid genotypes.

†*ura3* alleles harbor spontaneous LOF mutations isolated from 5FOA.

‡*ste4C958A* control allele to ensure there is no fitness effect of a synonymous PAM site substitution used in strain construction.

## Appendix C

### Preliminary findings on the evolutionary dynamics of gene-drives

I originally proposed a research direction very different from that contained in this dissertation. I intended to build CRISPR-based gene drives and use experimental evolution to characterize co-evolutionary dynamics of drive genes and host genomes. Although a simplified pilot experiment had indicated this direction would be possible at the time of my proposal, the experiments ultimately became untenable. In this appendix I have included a brief introduction to this work, the small amount of data derived from a pilot experiment, and a summary of the barriers I encountered that forced me to change research directions.

#### **Introduction**

Controlling disease vector populations is a major global challenge. Reduction and chemical destruction of breeding habitat, use of repellents, and distribution of mosquito nets have all helped lower malaria transmission. Genetic modification is a promising control strategy that has been shown to be effective at suppressing vector populations (Harris et al., 2012). An approach that takes advantage of the non-Mendelian inheritance of selfish genetic elements to drive beneficial transgenes through populations has seen a resurgence of interest due to the discovery and applications of CRISPR-Cas genes.

The description of CRISPR-Cas, or clustered regularly interspaced palindromic repeats and CRISPR-associated sequences (Doudna & Charpentier, 2014) has fundamentally changed the field of genetic engineering. Cas9 is a RNA-guided endonuclease whose target sequence is specified by a 20 base pair guide RNA. Because of its lengthy target sequence requirement (20bp) and its ability to generate double strand breaks in targeted sequences, Cas9 was quickly recognized for its potential as a drive element (DiCarlo, Chavez, Dietz, Esvelt, & Church, 2015; Gantz, Valentino & Bier, 2015). The idea is a simple adaptation of more traditional homing endonuclease drive; Cas9 would be integrated into the genome at a locus homologous to whatever site is being targeted, and when Cas9 is expressed, the resulting double strand break would be repaired through homologous repair. The advantage is that Cas9 does not need to be engineered or modified to recognize this target. It only needs to be co-expressed with a gRNA containing the target sequence motif. In such a way, Cas9 could be targeted anywhere in the genome and gene drives could be engineered with much less effort.

#### *Recent development of Cas9 Based Gene Drives*

Functional RNA-guided gene drives have been demonstrated in *S. cerevisiae* (DiCarlo et al., 2015), *Drosophila* (Gantz, Valentino & Bier, 2015), and *Anopheles* mosquitos (Gantz et al., 2015; Hammond et al., 2016). These studies demonstrated that Cas9 drive is highly efficient with 97-99% efficiencies (DiCarlo et al., 2015; Gantz, V. & Bier, 2015), and that Cas9 can drive the inheritance of refractory genes in a vector species (Gantz et al., 2015). Now that the last barriers to developing synthetic gene drives have seemingly been overcome, attention is called to large gaps in knowledge regarding the long-term behavior of gene drives. How do host genomes respond to drive? How

likely is it for host mutations to create an escape from drive? Is the mechanism of drive likely to be a dynamic trait subject to selection? We seek to begin addressing these questions empirically using experimental evolution in laboratory populations of *S. cerevisiae*.

### *Models of Gene Drive*

Most models of gene drive are determinate and derive predictions from evolutionary population genetics. A set of equations were introduced by Deredec (2008) and expanded by Unckless et al. (2015) to model the dynamics of endonuclease-based drive elements in evolving populations. In distilled form, this model derives drive allele frequency from inputs of starting frequency ( $q$ ), degree of dominance of drive alleles ( $h$ ), and conversion rate of drive alleles ( $c$ ), and the selective cost of the drive allele ( $s$ ).

$$q' = \frac{q^2(1-s) + qp[(1-c)(1-hs) + 2c(1-s)]}{\bar{w}}$$

This model predicts that drive alleles will fix when conversion rates are high and fitness costs are low. Conversely, drive alleles will be purged when conversion is weak and selection is strong. Here, we directly test both the underlying assumptions of gene drive models as well as their power to predict population dynamics over time.

### *Experimental Evolution and Sexual Reproduction*

Experimental evolution is an empirical approach to the study of evolution wherein populations are evolved in controlled environments in the laboratory. In our laboratory we use the budding yeast, *Saccharomyces cerevisiae* as a model due to its rapid life cycle, genomic simplicity, and amenability for genetic analysis. The automation we integrate into our approach allows for the high throughput evolution of hundreds of populations for thousands of generations period of months. Our experimental approach

also includes tracking changes in fitness and genotype over the course of adaptive evolution. We are able to use flow cytometry assays to quantify changes in fitness and employ whole genome sequencing to gain high-resolution insights into genetic changes underlying fitness gains.

Experimental evolution of yeast populations is an ideal approach to the study of gene drive. Quantification of selective costs and population size throughout experiments, along with precise tracking of allele frequencies will allow us to robustly test the predictions of gene drive models. Yeast cells are capable of sexual cycles, which is necessary for any gene drive element to spread. We have adapted a method first published in Burke et al. (2014) to force evolving yeast populations through cycles of meiosis and outcrossing in such a way as to maintain the high-throughput nature of our experimental approach while integrating sexual reproduction (Fig.2).

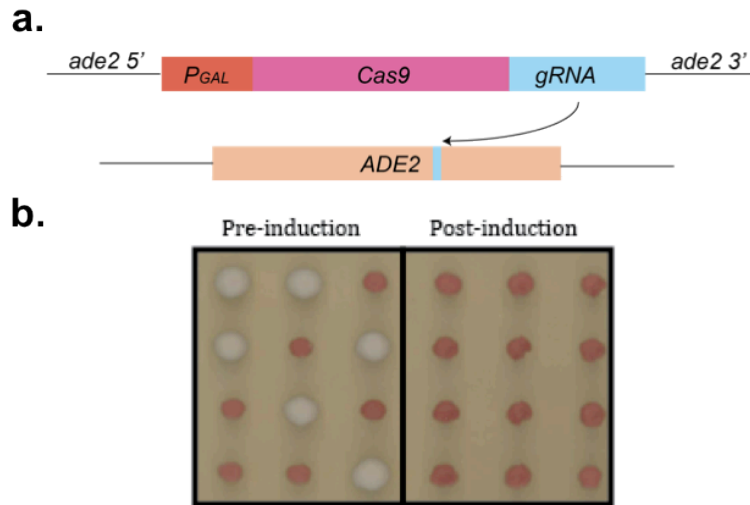
### **Experimental Approach & Preliminary Data**

Our experimental approach is to construct a synthetic gene drive and introduce it into laboratory populations. This approach presents two major challenges: (1) sexual cycles must be introduced into a high-throughput experimental evolution protocol that previously propagated asexual cultures and (2) a synthetic gene drive system with reporters that allow precise quantification in populations must be built. We have designed a protocol for the evolution of yeast with sexual cycles. We have engineered synthetic RNA guided gene drives in *S. cerevisiae*. We have piloted our experimental approach using low-resolution techniques and are now initiating a more robust evolution experiment complete with high-resolution analysis methods.

*Construction and validation of ADE2- and HIS3- targeted drive alleles*



We have designed a simple RNA-guided gene drive based on both current literature and preliminary results in our lab. The gene drive we have built is a knockout driver, in which the Cas9 drive allele fully replaces the allelic wild-type gene. This is opposed to generating an insertion of the drive element or acting in *trans* (Noble et al., 2016). The initial drive construct we designed was targeted to *ADE2*, a gene involved in adenine biosynthesis. We selected *ADE2* to enable red/white colony screening (*ade2* mutants accumulate a red pigment). We were able to successfully build and induce gene drive (**Fig. C-1**), however the *ade2* mutation imposed too high a fitness cost to be of experimental use, even when supplemented with excess adenine. We redesigned our drive construct to target *HIS3*, a gene involved in histidine biosynthesis. A *his3* knockout is completely recessive in a diploid, but a double knockout requires histidine supplementation in order to be viable. By titrating the amount of histidine available in experimental media, we can impose a gradient of fitness costs across experimental



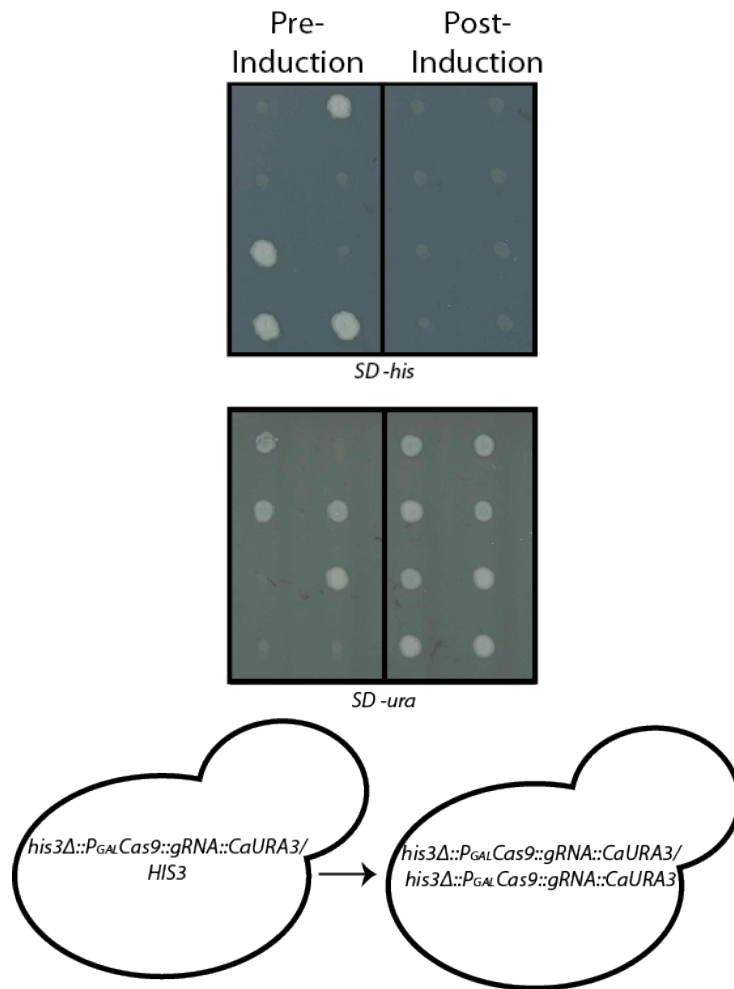
**Figure C-1.** ADE2 construct used in pilot experiment. A CRISPR-based drive allele was assembled at ADE2 for use in preliminary experiments. **a.** Cas9 under the control of a galactose inducible promoter at the *ade2* locus targeted to WT ADE2. **b.** Meiotic products of hemizogotes segregate 2:2 pre-induction and 4:0 post induction. *ade2* mutants are characterized by the accumulation of a metabolite that turns colonies pink and imposes a strong fitness defect (evidenced by small colony size).

populations. We first built our *HIS3*-targetted drive allele by linking *Cas9-gRNA* to a heterologous *Candida albicans URA3* marker (allele confers viability in media lacking uracil). **Fig. C-2** shows this allele segregating in meiotic divisions in pre-induction and post-induction diploids.

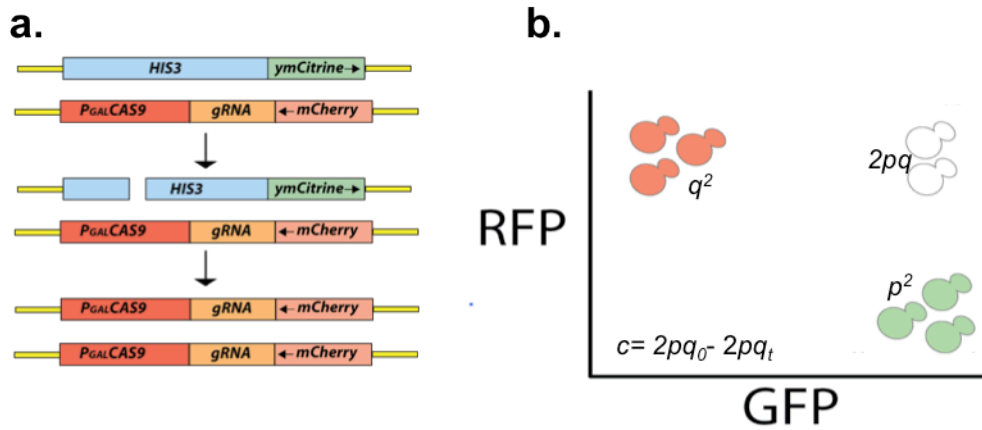
Our drive construct consists of a gal-inducible *Cas9* linked to a *HIS3*-targetted gRNA regulated by a pol III (*SNR52*) promoter. *Cas9-gRNA* is further linked to a constitutive mCherry fluorescent protein sequence that labels all cells containing at least one copy. This entire construct replaces *HIS3* coding sequence along with ~250 bp flanking sequence in both directions. To differentially label WT cells, we integrated a *ymCitrine* tag 250 bp downstream of the WT *HIS3* allele. When *Cas9p* cleaves the target sequence in *HIS3*, homology directed repair will copy the entire *Cas9-gRNA-mCherry* allele into the *HIS3* locus and fully replace *HIS3* and *ymCitrine*. Our fluorophore scheme will allow us to precisely quantify the frequency of drive homozygotes, WT homozygotes, and heterozygotes via flow cytometry and fluorescence-activated cells sorting (FACS) (**Fig. C-3**). Because *Cas9* is inducible, a FACS assay on a culture before and after induction will yield data on conversion frequency, or drive efficiency.

#### *Drive allele introduced into outcrossing cycles*

We will make use of the galactose induction system endogenous to *S. cerevisiae* to confine drive events to once per outcrossing cycle. When cells are exposed to media that is devoid of glucose and rich in galactose, a signaling cascade results in the activation of the transcription factor Gal4, which will bind to the Gal4 promoter upstream of *Cas9*. Our preliminary data indicate an induction time of 6 hours is sufficient to maximize conversion without effecting cell viability. Following induction, cultures are



**Figure C-2.** HIS3 drive. A Cas9-based drive allele was assembled at HIS3. Prior to induction, meiotic products of heterozygotes segregate 2:2 for histidine prototrophy (WT HIS3) and uracil prototrophy (*his3Δ::PGALCas9-gRNA-CaURA3*). Induction of the drive allele results in 4:0 segregation of histidine auxotrophy and uracil prototrophy.



**Figure C-3.** Gene constructs and fluorescent phenotypes. Fluorescent reporters will be utilized to distinguish genotypes. **a.** Dr and WT *HIS3* alleles are linked to an *mCherry* and *ymCitrine* cassette, respectively. Drive results from successful gene conversion following galactose induction. **b.** A two-color FACS assay will be used to distinguish genotypes throughout the experiment. *p* and *q* will be calculated to follow allele frequencies and conversion rate (*c*) will be estimated by subtracting post-induction heterozygote frequency ( $2pq_t$ ) from pre-induction frequency ( $2pq_0$ ).

immediately diluted into rich glucose media, which will instantly repress the galactose signaling pathway and turn Cas9 expression off.

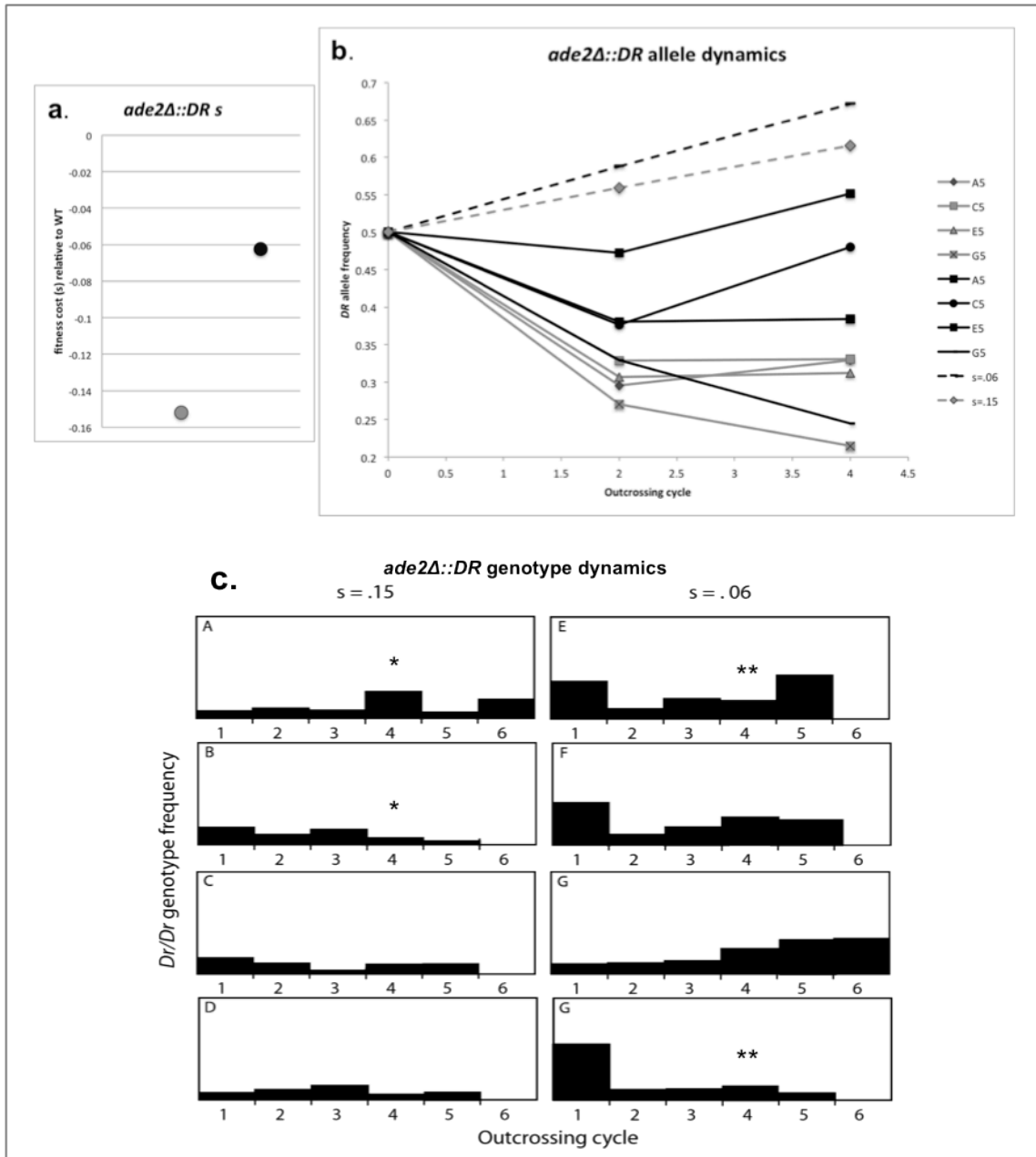
It is widely acknowledged that constitutive expression of an endonuclease, even one with specificity, will likely be deleterious to the transgenic organism. Most gene drives being designed seek to limit Cas9 endonuclease activity to one point in the life cycle of the target organism. In mosquitoes, this is usually during spermatogenesis (Gantz et al., 2015; Hammond et al., 2016). For this experiment, we have chosen to induce Cas9 expression following mating. This approach is akin to drive upon zygote formation, as opposed to gametogenesis. The reason for this is two-fold; 1) this is the assumption of the mathematical model we are directly testing (Unckless, Messer, & Clark, 2015), and 2) drive upon zygote formation represents the selective extreme that gene drives will likely face. It may be that confining drive to gametogenesis is a practical impossibility. The most recent CRISPR-based *Plasmodium* resistance construct designed in *Anopheles* mosquitoes drives upon gametogenesis and zygote formation, depending on whether the transgene is maternally or paternally inherited. Gantz *et al.* (2015) generated a somatically integrated drive locus that is driven by a germ line promoter. When a heterozygous male is mated to a wild type female all progeny are heterozygous, indicative of germ line drive in the male. When a heterozygous female is mated to a wild type male all progeny are mutant for the WT allele (either due to conversion itself, or NHEJ following endonuclease activity at the target site), indicating drive both in the female germ line and somatically in the zygote. Somatic drive is likely due to perdurance of Cas9p and gRNA molecules in the egg, which target the paternally inherited wild type

allele. Any construct that relies upon conversion in female germ line cells will likely encounter this issue, as no endogenous mechanism exists for clearing Cas9 complexes.

Lethal costs are associated with drive schemes that seek to impart a genetic load on populations in order to suppress population size through the targeted knockout of essential genes (Burt, 2003; Deredec, Burt, & Godfray, 2008). Conversely, the goal of those designing refractory gene drive systems is to achieve as low a selective cost as possible. Despite efforts to achieve this, most constructs exhibit some fitness effect or transgene instability over time (Franz et al., 2014; Moreira, Wang, Collins, & Jacobs-Lorena, 2004). These effects may be compounded in a CRISPR-drive system by the cost of *Cas9* expression and off-target effects. In our experiment we will vary the selective cost of homozygous drive mutants from neutral to lethal through histidine supplementation. Through this dynamic range we can assess the success different drive strategies without having to design complex drive alleles.

#### *Pilot Experiment*

Because we are proposing both a novel experimental evolution approach through the incorporation of sexual cycles, as well as a novel selective regime in the form of a drive allele, we first piloted the feasibility of this experiment using our original drive construct targeted to *ADE2*. This drive allele (*ade2Δ::Cas9-gRNA-CaURA3*, hereafter, *ade2Δ::DR*) is galactose inducible in the same manner of the *HIS3*-targeted drive allele. Successful drive was confirmed for post-induction cells that phenotyped as auxotrophs via tetrad dissections of meiotic products (**Fig. C-1a**). We then used competitive fitness assays versus a wild type reference to determine the selective cost of *ade2 Δ::DR* alleles (**Fig. C-4**). No fluorescent markers were incorporated into the genotypes used in this first



**Figure C-4.** Pilot experiment results. Four populations for each selection treatment were assayed for allele and genotype frequencies intermittently over 6 outcrossing cycles. **a.** Two different adenine supplementation treatments imposed a moderate and severe fitness cost for homozygous DR cells relative to WT. Black denotes  $s=0.06$  and grey denotes  $s=0.15$ . **b.** Allele frequencies were assayed every other cycle. Black lines,  $s=0.06$ ; grey lines,  $s=0.15$ , dashed lines, model-predicted frequencies. **c.** Black bar height represents *Dr/Dr* genotype frequency at each cycle. \*Fewer (*Dr/Dr*) than expected based on Hardy-Weinberg. (A;  $\chi^2=38.7$ ,  $p<0.001$ ; B;  $\chi^2=3.84$ ,  $p<0.05$ ). \*\* More (*Dr/Dr*) than expected (A;  $\chi^2=4.17$ ,  $p<0.05$ ; H;  $\chi^2=24.23$ ,  $p<0.001$ ).



experiment; therefore all allele and genotype frequencies were followed via plating assays for adenine auxotrophy.

The pilot evolution experiment was initiated by seeding 84 experimental populations with *ade2Δ::DR* at a frequency of 0.5. Experimental populations were subjected to two selective environments ( $s=0.06$ ,  $n=42$ ;  $s=0.15$ ,  $n=42$ ). Four populations from each treatment were assayed for allele frequencies every other outcrossing and for genotype frequencies every outcrossing using plating assays. The expectation of a moderate fitness cost allowing *ade2Δ::DR* allele increase and a severe fitness cost resulting in *ade2Δ::DR* allele loss was met by our preliminary analysis (**Fig. C-4b**). Furthermore, comparison to Hardy-Weinberg predictions at cycles for which both allele and genotype frequency data are available indicates that drive is overcoming selection in moderate-cost treatments while selection is overpowering drive in high-cost populations (**Fig. C-4c**). The low resolution of the data presented here was a result of the high percent error of plating assays ( $\approx 20\%$ ) along with low efficiency outcrossing. Despite the high error and noise in our preliminary data, it seems as though experimental populations differ markedly from model-predicted frequencies (dashed lines, **Fig. C-4b**). Additionally, the degree of inter-population variability suggests the predictive power of models may be lower than expected.

The described pilot experiment served as proof of principle that we can incorporate gene drive into high-throughput experimental evolution in sexual populations. However, these data also highlight the lack of quantitative power in our original experimental design. Only two selective treatments were possible due to the high

cost of *ade2* mutation. We anticipate that our *his3* selection will be more amenable to fitness-cost titration. Lack of resolution for measures of allele frequencies and genotype frequencies is overcome in the proposed experiment by the incorporation of fluorescent markers, which will provide high-resolution tracking of both variables and allow us to distinguish heterozygotes (**Fig. C-3**). Low outcrossing efficiency was due to the initial strategy of mating in liquid media. We have since optimized this step by switching to mating on filters, which concentrates cells in 2-dimensional space to facilitate efficient pheromone signaling and response. Despite these shortcomings, the trends gleaned from this experiment point to an interesting variability in drive allele dynamics that warrant further investigation. Here I propose to use the synthetic gene drive targeted to the *HIS3* locus to track the dynamics of this gene drive in real-time in evolving populations and to determine the extent to which the host genome resists the spread of the gene drive as a function of its selective cost.

### **Obstacles to completion of proposed work**

The project I proposed involved the construction of a gene drive allele at a new locus, *HIS3*, as well as the implementation of 2-color flow cytometry assays. While we were successful at constructing a gene drive at *HIS3*, we were ultimately unsuccessful in generating a construct for 2-color assays. My leading hypothesis for our inability to recover any mutants containing fluorescently labeled drive alleles is that we were unwittingly perturbing an essential intergenic region. Had we known this at the time, we may have tried to move the allele or create a completely synthetic transgene instead. However, we tried unsuccessfully to build our system for over a year.

As I moved into my fourth year it became obvious that the up-front investment in our gene-drive experiment was increasing while the payoff in terms of novelty was decreasing. Strain construction remained a seemingly insurmountable barrier. Additionally, other laboratories were beginning to do similar experiments with more extensive resources. It was at this point that I began focusing on several developing side projects, some of which have turned into the work presented in this dissertation.

## References

- Adamo, G. M. (2012). Evolution of copper tolerance in yeast cells. *Microbial Cell Factories*, 11(1)
- Agrawal, A. F., & Whitlock, M. C. (2011). Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics*, 187(2), 553-566.
- Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: What have we learned about the genetics of adaptation? *Trends in Ecology & Evolution*, 23(1), 26-32.
- Babu, M., Arnold, R., Bundalovic-Torma, C., Gagarinova, A., Wong, K. S., Kumar, A. (2014). Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genetics*, 10(2), e1004120.
- Bailey, S. F., Rodrigue, N., & Kassen, R. (2015). The effect of selection environment on the probability of parallel evolution. *Molecular Biology and Evolution*, 32(6), 1436-1448.
- Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D. et al. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461(7268), 1243.
- Barton, N. H., & De Cara, Maria Angeles Rodriguez. (2009). The evolution of strong reproductive isolation. *Evolution: International Journal of Organic Evolution*, 63(5), 1171-1190.
- Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K. et al. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature Methods*, 7(12), 1017.
- Beaulieu, J. M., Leitch, I. J., Patel, S., Pendharkar, A., & Knight, C. A. (2008). Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist*, 179(4), 975-986.
- BELL, M. A. (1987). Interacting evolutionary constraints in pelvic reduction of threespine sticklebacks, *Gasterosteus aculeatus* (Pisces, Gasterosteidae). *Biological Journal of the Linnean Society*, 31(4), 347-382.
- Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A. et al. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, 31(4), 872-888.
- Betancourt, A. J. (2009). Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics*, 181(4), 1535-1544.
- Bindewald, E., & Shapiro, B. A. (2006). RNA secondary structure prediction from sequence alignments using a network of k-nearest neighbor classifiers. *RNA (New York, N.Y.)*, 12(3), 342-352.

- Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. V., & Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. *Nature*, *494*(7436), 234.
- Bloom, J. S., Kotenko, I., Sadhu, M. J., Treusch, S., Albert, F. W., & Kruglyak, L. (2015). Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. *Nature Communications*, *6*, 8712.
- Breslow, D. K., Cameron, D. M., Collins, S. R., Schuldiner, M., Stewart-Ornstein, J., Newman, H. W. et al. (2008). A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nature Methods*, *5*(8), 711.
- Brewer, B. J., Payen, C., Di Rienzi, S. C., Higgins, M. M., Ong, G., Dunham, M. J. et al. (2015). Origin-dependent inverted-repeat amplification: Tests of a model for inverted DNA amplification. *PLoS Genetics*, *11*(12), e1005699.
- Bridgham, J. T., Ortlund, E. A., & Thornton, J. W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, *461*(7263), 515.
- Bull, J. J., Badgett, M. R., Wichman, H. A., Huelsenbeck, J. P., Hillis, D. M., Gulati, A. et al. (1997). Exceptional convergent evolution in a virus. *Genetics*, *147*(4), 1497-1507.
- Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1518), 921-928.
- Buskirk, S. W., Peace, , & Lang, G. I. (2017). Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(31), 8330-8335.
- Canadell, D., González, A., Casado, C., & Ariño, J. (2015). Functional interactions between potassium and phosphate homeostasis in *Saccharomyces cerevisiae*. *Molecular Microbiology*, *95*(3), 555-572.
- Carroll, M. W., Matthews, D. A., Hiscox, J. A., Elmore, M. J., Pollakis, G., Rambaut, A. et al. (2015). Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature*, *524*(7563), 97.
- Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Jr, Shapiro, M. D., Brady, S. D. et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science (New York, N.Y.)*, *327*(5963), 302-305.
- Chang, S., Lai, H., Tung, S., & Leu, J. (2013). Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genetics*, *9*(1), e1003232.
- Charles, J. S., & Petes, T. D. (2013). High-resolution mapping of spontaneous mitotic recombination hotspots on the 1.1 mb arm of yeast chromosome IV. *PLoS Genetics*, *9*(4), e1003434.

- Charlesworth, B. (1998). Adaptive evolution: The struggle for dominance. *Current Biology*, 8(14), R502-R504.
- Charlesworth, B., & Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genetics Research*, 74(3), 329-340.
- Chasnov, J. R. (2000). Mutation-selection balance, dominance and the maintenance of sex. *Genetics*, 156(3), 1419-1425.
- Chelo, I. M., & Teotónio, H. (2013). The opportunity for balancing selection in experimental populations of *Caenorhabditis elegans*. *Evolution: International Journal of Organic Evolution*, 67(1), 142-156.
- Chevin, L. (2013). Genetic constraints on adaptation to a changing environment. *Evolution: International Journal of Organic Evolution*, 67(3), 708-721.
- Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16), 2745-2747.
- Chou, H. H., Chiu, H. C., Delaney, N. F., Segre, D., & Marx, C. J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science (New York, N.Y.)*, 332(6034), 1190-1192.
- Chueca, L. J., Gómez-Moliner, B. J., Madeira, M. J., & Pfenninger, M. (2018). Molecular phylogeny of *Candidula* (geomitridae) land snails inferred from mitochondrial and nuclear markers reveals the polyphyly of the genus. *Molecular Phylogenetics and Evolution*, 118, 357-368.
- Clop, A., Vidal, O., & Amills, M. (2012). Copy number variation in the genomes of domestic animals. *Animal Genetics*, 43(5), 503-517.
- Codoñer, F. M., & Fares, M. A. (2008). Why should we care about molecular coevolution? *Evolutionary Bioinformatics*, 4, 117693430800400003.
- Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Jr, Dickson, M., Grimwood, J. et al. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science (New York, N.Y.)*, 307(5717), 1928-1933.
- Connallon, T., & Hall, M. D. (2018). Genetic constraints on adaptation: A theoretical primer for the genomics era. *Annals of the New York Academy of Sciences*, 1422(1), 65-87.
- Cooper, T. F., Rozen, D. E., & Lenski, (2003). Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 1072-1077.
- Cooper, Schneider, D., Blot, M., & Lenski, (2001). Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *Journal of Bacteriology*, 183(9), 2834-2841.

- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S. et al. (2010). The genetic landscape of a cell. *Science (New York, N.Y.)*, 327(5964), 425-431.
- Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G. et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science (New York, N.Y.)*, 353(6306), 10.1126/science.aaf1420.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A. et al. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.
- De Godoy, L. M., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Fröhlich, F. et al. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 455(7217), 1251-1254.
- Deatherage, D. E., Kepner, J. L., Bennett, A. F., Lenski, , & Barrick, J. E. (2017). Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, 114(10), E1904-E1912.
- Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology*, 3(10), e314.
- Deredec, A., Burt, A., & Godfray, H. C. (2008). The population genetics of using homing endonuclease genes in vector and pest management. *Genetics*, 179(4), 2013-2026.
- Deutschbauer, A. M., Jaramillo, D. F., Proctor, M., Kumm, J., Hillenmeyer, M. E., Davis, R. W. et al. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*, 169(4), 1915-1925.
- DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). RNA-guided gene drives can efficiently and reversibly bias inheritance in wild yeast. *bioRxiv*, , 013896.
- Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, 346(6213), 1258096.
- Dunham, M. J., Badrane, H., Ferea, T., Adams, J., Brown, P. O., Rosenzweig, F. et al. (2002). Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25), 16144-16149.
- Echenique, J. I. R., Kryazhimskiy, S., Ba, A. N. N., & Desai, M. M. (2019). Modular epistasis and the compensatory evolution of gene deletion mutants. *PLoS Genetics*, 15(2), e1007958.
- Engel, S. R., & Cherry, J. M. (2013). The new modern era of yeast genomics: Community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the *Saccharomyces* genome database. *Database*, 2013, bat012.

- Epstein, C. J. (1967). Cell size, nuclear content, and the development of polyploidy in the mammalian liver. *Proceedings of the National Academy of Sciences of the United States of America*, 57(2), 327-334.
- Ezov, T. K., Boger-Nadjar, E., Frenkel, Z., Katsperovski, I., Kemeny, S., Nevo, E. et al. (2006). Molecular-genetic biodiversity in a natural population of the yeast *Saccharomyces cerevisiae* from "evolution canyon": Microsatellite polymorphism, ploidy and controversial sexual status. *Genetics*, 174(3), 1455-1468.
- Ferretti, L., Weinreich, D., Tajima, F., & Achaz, G. (2018). Evolutionary constraints in fitness landscapes. *Heredity*, 121(5), 466.
- Fisher, K. J., Buskirk, S. W., Vignogna, R. C., Marad, D. A., & Lang, G. I. (2018). Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*. *PLoS Genetics*, 14(5), e1007396.
- Fisher, R. A. (1928). The possible modification of the response of the wild type to recurrent mutations. *The American Naturalist*, 62(679), 115-126.
- Fong, S. S., Joyce, A. R., & Palsson, B. Ø. (2005). Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Research*, 15(10), 1365-1372.
- Foot, A. D., Liu, Y., Thomas, G. W., Vinař, T., Alföldi, J., Deng, J. et al. (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics*, 47(3), 272.
- Fox, D. T., & Duronio, R. J. (2013). Endoreplication and polyploidy: Insights into development and disease. *Development (Cambridge, England)*, 140(1), 3-12.
- Franz, A. W., Sanchez-Vargas, I., Raban, R. R., Black IV, W. C., James, A. A., & Olson, K. E. (2014). Fitness impact and stability of a transgene conferring resistance to dengue-2 virus following introgression into a genetically diverse aedes aegypti strain. *PLoS Negl Trop Dis*, 8(5), e2833.
- Frenkel, E. M., Good, B. H., & Desai, M. M. (2014). The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. *Genetics*, 196(4), 1217-1226.
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., & Fink, G. R. (1999). Ploidy regulation of gene expression. *Science (New York, N.Y.)*, 285(5425), 251-254.
- Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science*, 348(6233), 442-444.
- Gantz, V. M., & Bier, E. (2015). Genome editing. the mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science (New York, N.Y.)*, 348(6233), 442-444.



- Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E. et al. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences of the United States of America*, 112(49), E6736-43.
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv Preprint arXiv:1207.3907*,
- Gerlinger, M., Horswell, S., Larkin, J., Rowan, A. J., Salm, M. P., Varela, I. et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics*, 46(3), 225.
- Gerstein, A., Kuzmin, A., & Otto, S. (2014). Loss-of-heterozygosity facilitates passage through Haldane's sieve for *Saccharomyces cerevisiae* undergoing adaptation. *Nature Communications*, 5, 3819.
- Gerstein, A. C., Chun, H. E., Grant, A., & Otto, S. P. (2006). Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genetics*, 2(9), e145.
- Gerstein, A. C., Cleathero, L., Mandegar, M., & Otto, S. (2011). Haploids adapt faster than diploids across a range of environments. *Journal of Evolutionary Biology*, 24(3), 531-540.
- Gerstein, A. C., Lim, H., Berman, J., & Hickman, M. A. (2017). Ploidy tug-of-war: Evolutionary and genetic environments influence the rate of ploidy drive in a human fungal pathogen. *Evolution*, 71(4), 1025-1038.
- Gerstein, A. C., & Otto, S. P. (2011). Cryptic fitness advantage: Diploids invade haploid populations despite lacking any apparent advantage as measured by standard fitness assays. *PloS One*, 6(12), e26599.
- Gerstein, A. C. (2013). Mutational effects depend on ploidy level: All else is not equal. *Biology Letters*, 9(1), 20120614.
- Gerstein, A. C., Fu, M. S., Mukaremera, L., Li, Z., Ormerod, K. L., Fraser, J. A. et al. (2015). Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. *mBio*, 6(5), e01340-15.
- Glazer, A. M., Cleves, P. A., Erickson, P. A., Lam, A. Y., & Miller, C. T. (2014). Parallel developmental genetic features underlie stickleback gill raker evolution. *EvoDevo*, 5(1), 19.
- Gloor, G. B., Martin, L. C., Wahl, L. M., & Dunn, S. D. (2005). Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. *Biochemistry*, 44(19), 7156-7165.
- Gong, L. I., Suchard, M. A., & Bloom, J. D. (2013). Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife*, 2, e00631.
- Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E., & Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678), 45.

- Gorter, F. A., Derks, M. F. L., van, d. H., Aarts, M. G. M., Zwaan, B. J., de Ridder, D. et al. (2017). Genomics of adaptation depends on the rate of environmental change in experimental yeast populations. *Molecular Biology and Evolution*, *34*(10), 2613-2626.
- Goudie, F., Allsopp, M. H., & Oldroyd, B. P. (2014). Selection on overdominant genes maintains heterozygosity along multiple chromosomes in a clonal lineage of honey bee. *Evolution*, *68*(1), 125-136.
- Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F., & Petren, K. (2004). Convergent evolution of darwin's finches caused by introgressive hybridization and selection. *Evolution*, *58*(7), 1588-1599.
- Gregory, T. R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews*, *76*(1), 65-101.
- Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A. et al. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet*, *4*(12), e1000303.
- Gresham, D., Usaite, R., Germann, S. M., Lisby, M., Botstein, D., & Regenberg, B. (2010). Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(43), 18551-18556.
- Hagen, D., & Gilbertson, L. (1972). Geographic variation and environmental selection in *Gasterosteus aculeatus* L. in the Pacific northwest, America. *Evolution*, *26*(1), 32-51.
- Haldane, J. (1924). A mathematical theory of natural and artificial selection, part I. *Transactions of the Cambridge Philosophical Society*, *23*, 19-41.
- Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D. et al. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology*, *34*(1), 78-83.
- Harari, Y., Ram, Y., Rappoport, N., Hadany, L., & Kupiec, M. (2018). Spontaneous changes in ploidy are common in yeast. *Current Biology*, *28*(6), 825-835. e4.
- Harris, A. F., McKemey, A. R., Nimmo, D., Curtis, Z., Black, I., Morgan, S. A. et al. (2012). Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nature Biotechnology*, *30*(9), 828-830.
- Hedrick, P. W. (2012). What is the evidence for heterozygote advantage selection? *Trends in Ecology & Evolution*, *27*(12), 698-704.
- Hillesland, K. L., Velicer, G. J., & Lenski, (2008). Experimental evolution of a microbial predator's ability to find prey. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1656), 459-467.

- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6(2), e1000862.
- Hong, J., & Gresham, D. (2014). Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genetics*, 10(1), e1004041.
- Hu, Y., Wu, Q., Ma, S., Ma, T., Shan, L., Wang, X. et al. (2017). Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5), 1081-1086.
- Huang, W., Richards, S., Carbone, M. A., Zhu, D., Anholt, R. R., Ayroles, J. F. et al. (2012). Epistasis dominates the genetic architecture of drosophila quantitative traits. *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), 15553-15559.
- Jaillon, O., Aury, J., Brunet, F., Petit, J., Stange-Thomann, N., Mauceli, E. et al. (2004). Genome duplication in the teleost fish tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature*, 431(7011), 946-957.
- Janos, L., & Korona, R. (2007). Epistatic buffering of fitness loss in yeast double deletion strains. *Nature Genetics*, 39(4), 550.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J. et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55.
- Keil, R. L., & Roeder, G. S. (1984). Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*. *Cell*, 39(2), 377-386.
- Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, 428(6983), 617-624.
- Khan, A. I., Dinh, D. M., Schneider, D., Lenski, , & Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science (New York, N.Y.)*, 332(6034), 1193-1196.
- Kim, Y., Koyutürk, M., Topkara, U., Grama, A., & Subramaniam, S. (2005). Inferring functional information from domain co-evolution. *Bioinformatics*, 22(1), 40-49.
- Korber, B. T., Farber, R. M., Wolpert, D. H., & Lapedes, A. S. (1993). Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: An information theoretic analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 90(15), 7176-7180.
- Kryazhimskiy, S., Bazykin, G. A., Plotkin, J., & Dushoff, J. (2008). Directionality in the evolution of influenza A haemagglutinin. *Proceedings of the Royal Society B: Biological Sciences*, 275(1650), 2455-2464.
- Kryazhimskiy, S., Dushoff, J., Bazykin, G. A., & Plotkin, J. B. (2011). Prevalence of epistasis in the evolution of influenza A surface proteins. *PLoS Genetics*, 7(2), e1001301.

- Kryazhimskiy, S., Rice, D. P., Jerison, E. R., & Desai, M. M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science (New York, N.Y.)*, *344*(6191), 1519-1522.
- Kvitek, D. J., & Sherlock, G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genetics*, *7*(4), e1002056.
- Kvitek, D. J., & Sherlock, G. (2013). Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genetics*, *9*(11), e1003972.
- LaFave, M. C., & Sekelsky, J. (2009). Mitotic recombination: Why? when? how? where? *PLoS Genetics*, *5*(3), e1000411.
- Lang, G. I., Murray, A. W., & Botstein, D. (2009). The cost of gene expression underlies a fitness trade-off in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5755-5760.
- Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D. et al. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, *500*(7464), 571-574.
- Lee, P. S., Greenwell, P. W., Dominska, M., Gawel, M., Hamilton, M., & Petes, T. D. (2009). A fine-structure map of spontaneous mitotic crossovers in the yeast *Saccharomyces cerevisiae*. *PLoS Genetics*, *5*(3), e1000410.
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A., & Fraser, A. G. (2006). Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genetics*, *38*(8), 896.
- Lenski, (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME Journal*, *11*(10), 2181.
- Lenski, , & Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(15), 6808-6814.
- Li, Y., Venkataram, S., Agarwala, A., Dunn, B., Petrov, D. A., Sherlock, G. et al. (2018). Hidden complexity of yeast adaptation under simple evolutionary conditions. *Current Biology*, *28*(4), 515-525. e6.
- Lieberman, T. D., Michel, J., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis Jr, M. R. et al. (2011). Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics*, *43*(12), 1275.
- Liti, G. (2015). The natural history of model organisms: The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *Elife*, *4*, e05835.

- Lockless, S. W., & Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science (New York, N.Y.)*, 286(5438), 295-299.
- Long, A., Liti, G., Luptak, A., & Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Reviews Genetics*, 16(10), 567.
- Lunzer, M., Golding, G. B., & Dean, A. M. (2010). Pervasive cryptic epistasis in molecular evolution. *PLoS Genetics*, 6(10), e1001162.
- Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B. et al. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 105(27), 9272-9277.
- Mable, B. K., & Otto, S. P. (2001). Masking and purging mutations following EMS treatment in haploid, diploid and tetraploid yeast (*Saccharomyces cerevisiae*). *Genetics Research*, 77(1), 9-26.
- Madlung, A. (2013). Polyploidy and its effect on evolutionary success: Old questions revisited with new tools. *Heredity*, 110(2), 99-104.
- Magwene, P. M., Kayikci, O., Granek, J. A., Reininga, J. M., Scholl, Z., & Murray, D. (2011). Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(5), 1987-1992.
- Mandegar, M. A., & Otto, S. P. (2007). Mitotic recombination counteracts the benefits of genetic segregation. *Proceedings of the Royal Society B: Biological Sciences*, 274(1615), 1301-1307.
- Manna, F., Martin, G., & Lenormand, T. (2011). Fitness landscapes: An alternative theory for the dominance of mutation. *Genetics*, 189(3), 923-937.
- Marad, D. A., Buskirk, S. W., & Lang, G. I. (2018). Altered access to beneficial mutations slows adaptation and biases fixed mutations in diploids. *Nature Ecology & Evolution*, 2(5), 882.
- Marcet-Houben, M., & Gabaldón, T. (2015). Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biology*, 13(8), e1002220.
- McCracken, K., Barger, C., Bulgarella, M., Johnson, K., Sonsthagen, S., Trucco, J. et al. (2009). Parallel evolution in the major haemoglobin genes of eight species of andean waterfowl. *Molecular Ecology*, 18(19), 3992-4005.
- Meyer, A., & Van de Peer, Y. (2005). From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *Bioessays*, 27(9), 937-945.
- Moodie, G., & Reimchen, T. (1976). Phenetic variation and habitat differences in *Gasterosteus* populations of the queen charlotte islands. *Systematic Zoology*, 25(1), 49-61.

- Moreira, L. A., Wang, J., Collins, F. H., & Jacobs-Lorena, M. (2004). Fitness of anopheline mosquitoes expressing transgenes that inhibit plasmodium development. *Genetics*, *166*(3), 1337-1341.
- Neverov, A. D., Kryazhimskiy, S., Plotkin, J. B., & Bazykin, G. A. (2015). Coordinated evolution of influenza A surface proteins. *PLoS Genetics*, *11*(8), e1005404.
- Newberry, M. G., McCandlish, D. M., & Plotkin, J. B. (2016). Assortative mating can impede or facilitate fixation of underdominant alleles. *Theoretical Population Biology*, *112*, 14-21.
- Noble, C., Min, J., Olejarz, J., Buchthal, J., Chavez, A., Smidler, A. L. et al. (2016). Daisy-chain gene drives for the alteration of local populations. *bioRxiv*, , 057307.
- O'quin, K. E., Hofmann, C. M., Hofmann, H. A., & Carleton, K. L. (2010). Parallel evolution of opsin gene expression in african cichlid fishes. *Molecular Biology and Evolution*, *27*(12), 2839-2854.
- Orr, H. A., & Betancourt, A. J. (2001). Haldane's sieve and adaptation from the standing genetic variation. *Genetics*, *157*(2), 875-884.
- Orr, H. A., & Otto, S. P. (1994). Does diploidy increase the rate of adaptation? *Genetics*, *136*(4), 1475-1480.
- Otto, S. P. (2007). The evolutionary consequences of polyploidy. *Cell*, *131*(3), 452-462.
- Parisod, C., Holderegger, R., & Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytologist*, *186*(1), 5-17.
- Parra-Olea, G., & Wake, D. B. (2001). Extreme morphological and ecological homoplasy in tropical salamanders. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(14), 7888-7891.
- Payen, C., Di Rienzi, S. C., Ong, G. T., Pogachar, J. L., Sanchez, J. C., Sunshine, A. B. et al. (2014). The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. *G3 (Bethesda, Md.)*, *4*(3), 399-409.
- Pelosi, L., Kuhn, L., Guetta, D., Garin, J., Geiselmann, J., Lenski et al. (2006). Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics*, *173*(4), 1851-1869.
- Phillips, P. C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*(11), 855.
- Prado, F., Cortes-Ledesma, F., Huertas, P., & Aguilera, A. (2003). Mitotic recombination in *Saccharomyces cerevisiae*. *Current Genetics*, *42*(4), 185-198.
- Press, M., Hall, A., Morton, E., & Queitsch, C. (2018). Substitutions are boring: Some arguments about parallel mutations and repetitive DNA.

- Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R. et al. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics*, 38(1), 107.
- Ratcliff, W. C., Herron, M. D., Howell, K., Pentz, J. T., Rosenzweig, F., & Travisano, M. (2013). Experimental evolution of an alternating uni- and multicellular life cycle in *Chlamydomonas reinhardtii*. *Nature Communications*, 4, 2742.
- Rosenfeld, L., Reddi, A. R., Leung, E., Aranda, K., Jensen, L. T., & Culotta, V. C. (2010). The effect of phosphate accumulation on metal ion homeostasis in *Saccharomyces cerevisiae*. *JBIC Journal of Biological Inorganic Chemistry*, 15(7), 1051-1062.
- Saleheen, D., Natarajan, P., Armean, I. M., Zhao, W., Rasheed, A., Khetarpal, S. A. et al. (2017). Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature*, 544(7649), 235.
- Sanchez, M. R., Miller, A. W., Liachko, I., Sunshine, A. B., Lynch, B., Huang, M. et al. (2017). Differential paralog divergence modulates genome evolution across yeast species. *PLoS Genetics*, 13(2), e1006585.
- Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S., & Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature*, 440(7082), 341-345.
- Schürmann, T., & Grassberger, P. (1996). Entropy estimation of symbol sequences. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 6(3), 414-427.
- Sellis, D., Callahan, B. J., Petrov, D. A., & Messer, P. W. (2011). Heterozygote advantage as a natural consequence of adaptation in diploids. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51), 20666-20671.
- Sellis, D., Kvitek, D. J., Dunn, B., Sherlock, G., & Petrov, D. A. (2016). Heterozygote advantage is a common outcome of adaptation in *Saccharomyces cerevisiae*. *Genetics*, 203(3), 1401-1413.
- Selmecki, A. M., Maruvka, Y. E., Richmond, P. A., Guillet, M., Shores, N., Sorenson, A. L. et al. (2015). Polyploidy can drive rapid adaptation in yeast. *Nature*, 519(7543), 349-352.
- Sezmis, A. L., Malerba, M. E., Marshall, D. J., & McDonald, M. J. (2018). Beneficial mutations from evolution experiments increase rates of growth and fermentation. *Journal of Molecular Evolution*, , 1-7.
- Shah, P., McCandlish, D. M., & Plotkin, J. B. (2015). Contingency and entrenchment in protein evolution under purifying selection. *Proceedings of the National Academy of Sciences of the United States of America*, 112(25), E3226-35.
- Shapiro, B., Rambaut, A., Pybus, O. G., & Holmes, E. C. (2006). A phylogenetic method for detecting positive epistasis in gene sequences and its application to RNA virus evolution. *Molecular Biology and Evolution*, 23(9), 1724-1730.

- Sharp, N. P., Sandell, L., James, C. G., & Otto, S. P. (2018). The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(22), E5046-E5055.
- Smukowski Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., & Dunham, M. J. (2017). Loss of heterozygosity drives adaptation in hybrid yeast. *Molecular Biology and Evolution*, *34*(7), 1596-1612.
- Solopova, A., Bachmann, H., Teusink, B., Kok, J., Neves, A. R., & Kuipers, O. P. (2012). A specific mutation in the promoter region of the silent *cel* cluster accounts for the appearance of lactose-utilizing lactococcus lactis MG1363. *Applied and Environmental Microbiology*, AEM. 00455-12.
- Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science*, *25*(7), 1204-1218.
- Steiner, C. C., Römpler, H., Boettger, L. M., Schöneberg, T., & Hoekstra, H. E. (2008). The genetic basis of phenotypic convergence in beach mice: Similar pigment patterns but different genes. *Molecular Biology and Evolution*, *26*(1), 35-45.
- Stoltzfus, A., & McCandlish, D. M. (2017). Mutational biases influence parallel adaptation. *Molecular Biology and Evolution*, *34*(9), 2163-2172.
- Storz, J. F. (2016). Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics*, *17*(4), 239.
- Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719.
- Sunshine, A. B., Payen, C., Ong, G. T., Liachko, I., Tan, K. M., & Dunham, M. J. (2015). The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biology*, *13*(5), e1002155.
- Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A. et al. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nature Genetics*, *43*(7), 656.
- Szulkin, M., Bierne, N., & David, P. (2010). Heterozygosity-fitness correlations: A time for reappraisal. *Evolution: International Journal of Organic Evolution*, *64*(5), 1202-1217.
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., & Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science (New York, N.Y.)*, *320*(5875), 486-488.
- Tenaillon, O., Rodriguez-Verdugo, A., Gaut, R. L., McDonald, P., Bennett, A. F., Long, A. D. et al. (2012). The molecular diversity of adaptive convergence. *Science (New York, N.Y.)*, *335*(6067), 457-461.



- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178-192.
- Tong, A. H., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X. et al. (2004). Global mapping of the yeast genetic interaction network. *Science (New York, N.Y.)*, *303*(5659), 808-813.
- Toprak, E., Veres, A., Michel, J., Chait, R., Hartl, D. L., & Kishony, R. (2012). Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature Genetics*, *44*(1), 101.
- Unckless, R., Messer, P., & Clark, A. (2015). Theoretical consequences of the mutagenic chain reaction for manipulating natural populations. *bioRxiv*, , 018986.
- Van de Peer, Y., Maere, S., & Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, *10*(10), 725-732.
- Van Ditmarsch, D., Boyle, K. E., Sakhtah, H., Oyler, J. E., Nadell, C. D., Déziel, É et al. (2013). Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell Reports*, *4*(4), 697-708.
- Van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, *6*(10), 767.
- Venkataram, S., Dunn, B., Li, Y., Agarwala, A., Chang, J., Ebel, E. R. et al. (2016). Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell*, *166*(6), 1585-1596. e22.
- Voordeckers, K., Kominek, J., Das, A., Espinosa-Cantú, A., De Maeyer, D., Arslan, A. et al. (2015). Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genetics*, *11*(11), e1005635.
- Weinreich, D. M., Delaney, N. F., Depristo, M. A., & Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science (New York, N.Y.)*, *312*(5770), 111-114.
- Whiteway, M., Hougan, L., Dignard, D., Thomas, D. Y., Bell, L., Saari, G. C. et al. (1989). The *STE4* and *STE18* genes of yeast encode potential  $\beta$  and  $\gamma$  subunits of the mating factor receptor-coupled G protein. *Cell*, *56*(3), 467-477.
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.
- Wilkening, S., Lin, G., Fritsch, E. S., Tekkedil, M. M., Anders, S., Kuehn, R. et al. (2014). An evaluation of high-throughput approaches to QTL mapping in *Saccharomyces cerevisiae*. *Genetics*, *196*(3), 853-865.
- Wilkens, H., & Strecker, U. (2003). Convergent evolution of the cavefish *astyanax* (characidae, teleostei): Genetic evidence from reduced eye-size and pigmentation. *Biological Journal of the Linnean Society*, *80*(4), 545-554.

- Wiser, M. J., Ribeck, N., & Lenski, (2013). Long-term dynamics of adaptation in asexual populations. *Science (New York, N.Y.)*, 342(6164), 1364-1367.
- Wolfe, K. H. (2015). Origin of the yeast whole-genome duplication. *PLoS Biology*, 13(8), e1002221.
- Wolfe, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387(6634), 708.
- Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8(3), 206.
- Wu, Z., Milne, R. I., Chen, C., Liu, J., Wang, H., & Li, D. (2015). Ancestral state reconstruction reveals rampant homoplasy of diagnostic morphological characters in urticaceae, conflicting with current classification schemes. *PLoS One*, 10(11), e0141821.
- Xie, K. T., Wang, G., Thompson, A. C., Wucherpfennig, J. I., Reimchen, T. E., MacColl, A. D. C. et al. (2019). DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science (New York, N.Y.)*, 363(6422), 81-84.
- Yeaman, S., Gerstein, A. C., Hodgins, K. A., & Whitlock, M. C. (2018). Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genetics*, 14(10), e1007717.
- Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C. et al. (2016). Convergent local adaptation to climate in distantly related conifers. *Science (New York, N.Y.)*, 353(6306), 1431-1433.
- Zeyl, C., Vanderford, T., & Carter, M. (2003). An evolutionary advantage of haploidy in large yeast populations. *Science (New York, N.Y.)*, 299(5606), 555-558.
- Zhang, H., Zeidler, A. F., Song, W., Puccia, C. M., Malc, E., Greenwell, P. W. et al. (2013). Gene copy-number variation in haploid and diploid strains of the yeast *Saccharomyces cerevisiae*. *Genetics*, 193(3), 785-801.
- Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M., & Andolfatto, P. (2012). Parallel molecular evolution in an herbivore community. *Science (New York, N.Y.)*, 337(6102), 1634-1637.
- Żmieńko, A., Samelak, A., Kozłowski, P., & Figlerowicz, M. (2014). Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics*, 127(1), 1-18.
- Zörgö, E., Chwialkowska, K., Gjuvslund, A. B., Garré, E., Sunnerhagen, P., Liti, G. et al. (2013). Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genetics*, 9(3), e1003388.

## Curriculum Vitae

### Kaitlin J. Fisher

PhD Candidate  
Laboratory of Greg Lang  
Department of Biological Sciences  
Lehigh University  
111 Research Dr.  
Bethlehem, PA 18015  
Kjf412@lehigh.edu

### Education

Lehigh University, Bethlehem, PA (2014 – May 2019)

Ph.D. Integrative Biology

Thesis: Parallelism, constraint, and functional genome evolution in experimentally evolving populations of *Saccharomyces cerevisiae*.

Advisor: Greg Lang

Thesis committee: Michael Kuchka, Amber Rice, Molly Burke

Long Island University – Post, Brookville, NY (2010 - 2013)

M.S. Biology

Thesis: Molecular characterization of spectral sensitivity in the convict cichlid, *Amatitlania nigrofasciatus*

Advisor: Matt Draud

Thesis committee: Scott Santagata, Ted Brummel

Long Island University – Post, Brookville, NY (2006 – 2010)

B.S. Biology

### Awards and Honors

Marjorie M. Nemes Fellow, 2018

Society for the Study of Evolution Travel Grant, 2017

Marjorie M. Nemes Fellow, 2016

LIU Biology Department Award for Excellence in Graduate Research, 2012

LIU Biology Department Award for Excellence in Undergraduate Research, 2010

### Research Interests

Molecular adaptation

Fitness conflicts

Functional genome evolution

### Publications

\*Indicates co-first authorship

**Fisher K.**, Kryazhimsky, S., Lang, G. Leveraging parallelism to detect beneficial mutations and pairwise epistasis in experimental evolution. *In press. Philosophical Transactions of the Royal Society B.*

**Fisher K.** \*, Buskirk S. \*, Vignogna R., Marad D., Lang G. (2018) Adaptive genome duplication affects patterns of molecular evolution in *Saccharomyces cerevisiae*. *PLoS Genet* 14(5): e1007396.

**Fisher, K.**, Lang, G. (2016) *Invited Review*; Experimental evolution in fungi: an untapped resource. *Fungal Genetics and Biology* 94 (2016): 88-94.

**Fisher, K.**, Schrey, A., Recupero, D., Draud, M. (2015) Molecular evidence of long wavelength spectral sensitivity in the reverse sexually dichromatic Convict Cichlid (*Amatitlania nigrofasciata*) *Copeia*. 103(3), 546-551

**Fisher, K.**, Guilfoyle, K., Hatch, K. (2013) Stress induced by toe-clipping in Cane Toads (*Rhinella marina*). *Copeia*. 2013(3), 539-542

### **Professional Presentations**

**Fisher K.**, Lang, G. 2018. Haploinsufficiency at a single locus constrains chromosome evolution: A case study in experimental evolution. Academy of Natural Sciences, Drexel University, Philadelphia, PA.

**Fisher K.**, Buskirk S, Lang G. 2017. Adaptive genome duplication affects patterns of molecular evolution in *S. cerevisiae*. 15<sup>th</sup> Annual Lehigh Valley Ecology and Evolution Symposium. Cedar Crest College, Allentown PA.

**Fisher K.**, Buskirk S, Lang G. 2017. Intralocus sexual conflict arises in experimentally evolving populations of *S. cerevisiae*. Evolution Conference. Portland, Oregon.

**Fisher K.**, Buskirk S, Lang G. 2017. Evidence of intralocus sexual conflict in experimentally evolving populations of *S. cerevisiae*. EPiC – Evolution in Philadelphia Conference. University of Pennsylvania, Philadelphia, PA. **Best Student Talk, 2<sup>nd</sup> Place**

**Fisher K.**, Buskirk S, Lang G. 2017. Evidence of intralocus sexual conflict in experimentally evolved populations of *S. cerevisiae*. 14<sup>th</sup> Annual Lehigh Valley Ecology and Evolution Symposium. DeSales University, Center Valley PA.

**Fisher K.** 2017. Studying fitness conflicts using experimental evolution in *S. cerevisiae*. Lehigh University Biological Sciences Colloquium Seminar Series. Bethlehem, PA.

**Fisher K.**, Lang G. 2016. Dynamics and molecular evolution of gene drive; experimental evolution in budding yeast as a model. The Allied Genetics Conference, Orlando, FL.

**Fisher K.**, Lang G. 2016. Dynamics and molecular evolution of gene drive; experimental evolution in budding yeast as a model. 13<sup>th</sup> Annual Lehigh Valley Ecology and Evolution Symposium. Lehigh University, Bethlehem PA.

**Fisher K.**, Lang, G. 2015. How predictable are evolutionary responses to genomic conflict? 12<sup>th</sup> Annual Lehigh Valley Ecology and Evolution Symposium. Muhlenberg College, Allentown PA.

**Fisher K.**, Schrey A, Recupero D, Draud M. 2014. Long wavelength spectral sensitivity and sexual dichromatism in the Convict Cichlid (*Amatitlania nigrofasciata*). Animal Behavior Society 51<sup>st</sup> Annual Meeting. Princeton, NJ

**Fisher K.**, Schrey A, Recupero D, Draud M. 2014. Long wavelength spectral sensitivity in the Convict Cichlid (*Amatitlania nigrofasciata*) Armstrong Atlantic State University. Savannah, GA.

**Fisher K, Schrey A, Recupero D, Draud M. 2013.** Reverse sexual dichromatism and spectral sensitivity in Convict Cichlids (*Amatitlania nigrofasciata*). Long Island University Biology Seminar, Brookville, NY.

**Fisher K, Guilfoyle K, Hatch K. 2013.** Toe-clipping and ethical concerns. Long Island University Biology Seminar, Brookville, NY.

**Fisher K, Guilfoyle K, Hatch K. 2010.** Does toe-clipping cause Cane Toads (*Rhinella marina*) excessive stress? Annual Meeting of the American Society of Ichthyologists and Herpetologists, Providence, RI.

### **Professional development**

Cold Spring Harbor Yeast Genetics and Genomics Course

Attended as student – 2015

Attended as TA – 2018

### **Teaching Experience**

\*Course evaluations available

#### **Teaching Assistant, Biological Sciences Department, Lehigh University, 2014-2019\***

Assisted with and instructed various laboratory and lecture courses.

Duties include lab creation, lab prep, lecture prep, lecturing, grading, and one on one student support.

Microbiology

Evolution

Genetics lab

Cell and Molecular Biology lab

#### **Adjunct Instructor, Foundations of Biology 1, LIU-Post, 2013 – 2014\***

Lectured for introductory course offered to non-science majors. Designed traditional lectures with online components and hands-on laboratories.

#### **Adjunct Instructor, Introduction to Biology 2, LIU- Post, 2013\***

Lectured for the second half of a two part introductory course for biology majors. Designed traditional lectures with online components and hands-on laboratories.

#### **Adjunct Instructor, biology major and non-major laboratories, LIU Post, 2011-2013\***

Instructed and supervised laboratories for both non-major courses and biology major courses

#### **Teaching Assistant, Biology and Anatomy and Physiology laboratories, LIU Post, 2010-2012**

Assisted instructor with laboratory prep and dissections

### **Service**

Member of Stem-Valley Mentoring Coalition- 2015 –present

Served as a science fair project mentor for a 7<sup>th</sup> grade biology classroom.

Served as a judge at the Lehigh Valley Science Fair.

Participated in various outreach events throughout the Lehigh Valley.

DNA Day Volunteer, 2017

Vice President of the Biological Organization of Graduate Students (BOGS) – 2017, 2018

Member of the Lehigh Graduate Student Senate – 2017, 2018

**Professional Affiliations**

Society for the Study of Evolution, 2017- present

American Society for Microbiology, 2014-present

Genetics Society of America, 2015- present

**References**

Greg Lang, PhD

[glang@lehigh.edu](mailto:glang@lehigh.edu)

Assistant Professor

Biological Sciences

Lehigh University

Bethlehem, PA

Amber Rice, PhD

[amr511@lehigh.edu](mailto:amr511@lehigh.edu)

Assistant Professor

Biological Sciences

Lehigh University

Bethlehem, PA

Sergey Kryazhimskiy

[skryazhi@ucsd.edu](mailto:skryazhi@ucsd.edu)

Assistant Professor

Division of Biological Sciences

University of California, San Diego

San Diego, CA