

EXPERIMENTS OF SEXUAL SELECTION AND PLOIDY-SPECIFIC EFFECTS OF
MUTATIONS IN YEAST

by

Nathalie Linnea Helene Sandell

B.Sc., Uppsala University, 2013

M.Sc., University of Montpellier and University of Groningen, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2021

© Nathalie Linnea Helene Sandell, 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Experiments of sexual selection and ploidy-specific effects of mutations in yeast

submitted by Nathalie Linnea Helene Sandell in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Zoology

Examining Committee:

Sarah P. Otto, Zoology, UBC

Supervisor

Darren Irwin, Zoology, UBC

Supervisory Committee Member

Dolph Schluter, Zoology, UBC

University Examiner

Peter Stirling, Medical genetics, UBC

University Examiner

Additional Supervisory Committee Members:

LeAnn Howe, Biochemistry, UBC

Supervisory Committee Member

Michael Doebeli, Zoology and Mathematics, UBC

Supervisory Committee Member

Abstract

For organisms to genetically adapt, they need to have or acquire mutations that affect their survival and reproductive success. The effects of mutations depend not only on the abiotic factors extrinsic to the organism, but also its biotic interactions as well as the intrinsic state of the organism. Preexisting mutations in the genome will either favor or prevent certain adaptive paths. This thesis contributes knowledge of two aspects of life cycle evolution using the budding yeast *Saccharomyces cerevisiae*. My evolutionary experiment of phenotypic responses to sexual selection showed that the two mating types of yeast, though often considered indistinguishable, respond differently to increased amounts of mating competition. These results have implications for our understanding of mating system evolution. In a bioinformatic project, I found that changes to the molecular function of proteins do not directly correspond to fitness differences among lines with multiple known mutations in the lab. These results affirm that mutations in proteins act on a system of molecular reactions which determine fitness, and caution against assuming a direct relationship between effects on molecular function and fitness. Finally, I conducted an experiment to measure the fitness effects of mutations in haploids versus diploids. The experimental method used to produce the three genotypic states led to large changes in fitness, impeding our investigation. Most notably I observed a large loss in respiratory function among my experimental replicates that correlated strongly with the presence of the wildtype *TIDI/RDH54* gene. While unable to answer the question originally stated, the experimental lines produced are a valuable resource for investigations into how the nuclear recombination machinery may increase the likelihood of mitochondrial mutations. Together, these results illustrate the utility of laboratory experiments to both answer and pose new evolutionary questions about such fundamental phenomena as the evolution of sex and the fitness effect of mutations.

Lay Summary

I did an evolutionary experiment where I had the two mating types of budding yeast (the equivalent of sperm and egg) compete for mating opportunities. I found that one mating type had a much stronger response to mating competition than the other type.

Secondly, I explored a popular computer tool that is often used to predict disease-causing variants in medicine and is assumed to apply in other species. I found that it fails to predict growth in yeast and in a green alga. This is likely because protein functionality does not directly translate to growth rates in the lab.

Thirdly, yeast are most often found with two copies of their genome and I wanted to compare how yeast grew with one or two copies of a mutated genome. The yeast changed during the experiment so that I could not answer this question.

Preface

A version of Chapter 2 is in preparation for submission in collaboration with N.P. Sharp and S.P. Otto. N.P. Sharp conceived of the original project, and I conceived of and carried out experiment and subsequent assays of the lines to determine their evolutionary responses to sexual selection. I performed laboratory work with assistance from S. Koenig, M. Stasiuk, C. Hsu, K. Somal, and B. Wiley. I analyzed data gathered for all sections, with advice from N.P. Sharp and S.P. Otto. I prepared the initial draft of the manuscript. N.P. Sharp and S.P. Otto contributed revisions to the manuscript.

A version of Chapter 3 is in preparation for submission in collaboration with N.P. Sharp and S.P. Otto. N.P. Sharp suggested the original project, and I conceived of and carried out the coding of protein variants from nucleotide variant. I also gathered and analyzed the data. I performed computational work on the departmental cluster with advice from A. Blachford. I analyzed data gathered for all sections, with advice from N.P. Sharp and S.P. Otto. I prepared the initial draft of the manuscript. N.P. Sharp and S.P. Otto contributed revisions to the manuscript.

A version of Chapter 4 is in preparation for submission in collaboration with N.P. Sharp, S. Koenig, M. Stasiuk, and S.P. Otto. I conceived of the original project. N.P. Sharp conceived of the broad experimental design, with advice from S. Koenig. S. Koenig designed and lead the knockout and plasmid transformations and subsequent phenotypic assays. I conceived of the mating of lines, as well as the growth rate assays. I and M. Stasiuk troubleshot and conducted the flow cytometric assays of the lines. All laboratory work was conducted with assistance from S. Koenig, M. Stasiuk, C. Hsu, K. Somal, and B. Wiley. I analyzed data gathered for all sections, with advice from N.P. Sharp and S.P. Otto. I prepared the initial draft of the manuscript. N.P. Sharp and S.P. Otto contributed revisions to the manuscript.

Table of Contents

Abstract.....	iii
Lay Summary	iv
Preface.....	v
Table of Contents	vi
List of Tables	xi
List of Figures.....	xii
Acknowledgements	xiii
Dedication	xiv
Chapter 1: Introduction	1
1.1 Experiments on sexual selection.....	3
1.2 Linking genotype to fitness.....	6
1.3 Mutational effects matter	8
Chapter 2: Evolutionary responses to increased sexual selection in yeast	12
2.1 Introduction.....	12
2.1.1 Yeast as a system to study sexual selection	13
2.1.2 Can yeast adapt to their mating system?.....	15
2.2 Methods.....	17
2.2.1 Strains	18
2.2.2 Media	18
2.2.3 Evolution experiment.....	19
2.2.3.1 Mating.....	19
2.2.3.2 Sporulation	20

2.2.3.3	Ascus digestion and germination	20
2.2.4	Assays	21
2.2.4.1	Mating rate assays.....	21
2.2.4.2	Sporulation rate assays.....	24
2.2.4.3	Growth-rate assays.....	24
2.2.4.4	Flow cytometry	25
2.2.4.5	PCR of <i>MAT</i> locus	25
2.2.4.6	Phenotype assays	26
2.2.5	Correlations between traits	26
2.3	Results.....	27
2.3.1	Mating rate	27
2.3.2	Cell size and genome content	29
2.3.3	Phenotype and mating type assay	31
2.3.4	Growth rate	32
2.3.5	Sporulation rate.....	33
2.3.6	Correlations between traits	34
2.4	Discussion.....	35
2.4.1	Mating-type specific adaptation.....	36
2.4.2	Response in other traits.....	38
2.4.3	Haploids with heterozygous mating type escape costs of mating and meiosis.....	39
2.5	Conclusion	40
Chapter 3: From function to fitness		41
3.1	Introduction.....	41

3.2	Background.....	48
3.3	Methods.....	52
3.3.1	Data processing.....	52
3.3.1.1	<i>S. cerevisiae</i> dataset 1	52
3.3.1.2	<i>S. cerevisiae</i> dataset 2	54
3.3.1.3	<i>C. reinhardtii</i> dataset	55
3.3.2	Running PROVEAN.....	58
3.3.3	Statistical analyses	59
3.3.3.1	<i>S. cerevisiae</i> dataset 1	60
3.3.3.2	<i>S. cerevisiae</i> dataset 2	60
3.3.3.3	<i>C. reinhardtii</i> dataset	61
3.4	Results.....	61
3.4.1	<i>S. cerevisiae</i> dataset 1	61
3.4.2	<i>S. cerevisiae</i> dataset 2	63
3.4.3	<i>C. reinhardtii</i> dataset	64
3.4.4	Differences in the number of clusters and supporting sequences between species	65
3.4.5	Scores for protein changes between laboratory strains and natural isolates.....	65
3.5	Discussion.....	68
3.6	Conclusion	72
	Chapter 4: Ploidy specific effects of mutations	73
4.1	Introduction.....	73
4.1.1	Ploidy specific effects of mutations.....	73

4.1.2	Dominance	74
4.1.3	Our contribution.....	76
4.1.4	Justification for method	76
4.2	Methods.....	78
4.2.1	Strain creation.....	78
4.2.2	Transformation protocol	78
4.2.2.1	Step 1: Deletion of the <i>STE4</i> locus	80
4.2.2.2	Step 2: Deletion of the <i>MAT</i> locus.....	81
4.2.2.3	Step 3: MA to KO comparison	82
4.2.2.4	Step 4: Plasmid transformation of the haploid lines	83
4.2.2.5	Step 5: Mating.....	84
4.2.2.6	Step 6: Fitness and ploidy assays.....	85
4.3	Results.....	85
4.3.1	Effect of <i>MAT</i> and <i>STE4</i> knockout.....	86
4.3.2	High frequency of petites following plasmid transformation.....	87
4.3.3	No correlation of growth rates across data sets	88
4.3.4	Inferences of mutational effects across haploid, heterozygote, and homozygote yeast	89
4.4	Discussion.....	91
4.4.1	Deletion of <i>RDH54</i> protects against petite-formation	93
4.5	Conclusion	94
Chapter 5: Conclusion		95
5.1	Sex, drugs, and experimental evolution.....	95

5.2	Fitness effect of mutations	99
References		105
Appendices		116
Appendix A.....		116
A.1	Selective markers	116
A.2	Dilutions for mating	116
A.3	Flow cytometry	117
A.4	Phenotyping	117
Appendix B.....		122

List of Tables

Table 2.1 Media used in this experiment.....	18
Table 2.2 Mating ratios during the experiment and in post-experiment assays.....	20
Table 3.1 The use of PROVEAN is widespread in ecology and evolutionary studies.....	44
Table 3.2 Number of protein-coding mutations for MA lines.....	56
Table 3.3 Summary of protein mutations in (a) Sc1, (b) Sc2, (c) Cr1 dataset.....	59
Table 4.1 Kendall's correlation coefficient of mean growth rate of MA lines with same petite status in DS1 and DS2	89

List of Figures

Figure 1.1 Dominance of wildtype phenotype is an inherent property of multi-step enzyme pathways.	9
Figure 2.1 Schematic of evolutionary experiment.....	17
Figure 2.2 Adaptation to breeding system	27
Figure 2.3 Cross-couple mating rates across assay ratio	29
Figure 2.4 Flow cytometry of evolved lines	31
Figure 2.5 Growth rate increased in all selective media.....	33
Figure 2.6 Sporulation rate of lines.....	34
Figure 3.1 Conceptual figure of project.....	50
Figure 3.2 PROVEAN prediction adds little explanatory power to growth rate differences among MA lines.....	62
Figure 4.1 Transformations done to lines in this experiment	80
Figure 4.2 Comparison of growth rates before and after knockout of the <i>STE4</i> and <i>MAT</i> loci ...	87
Figure 4.3 Petiteness of the line strongly predicts maximum slope of growth curve in both datasets.	88
Figure 4.4 We find no correlation between the growth rate of the MA line in its heterozygote and homozygote form.....	90

Acknowledgements

Many thanks to my supervisor Dr. Sarah P. Otto and my supervisory committee, Dr. LeAnn Howe, Dr. Michael Doebeli and Dr. Darren Irwin. Special thanks to Dr. Nathaniel Sharp who gave me projects to pursue when my own had failed.

Thank you to all the members of the wet lab. Thanks to Dr. Jasmine Ono and Ana Kuzmin for showing me around when I first started and called me out when I was sloppy. To Dr. Stephan Koenig and Matt Stasiuk who helped troubleshoot and made my last chapter bearable, thank you. I have also been fortunate to have worked with and learned from: Lesley Miller, Chris James, Christina Hsu, Kismet Somal, and Bryn Wiley. I'm also thankful to the past and current members and visiting scholars of the Otto lab, whose research continue to inspire me. In particular, the mentorship and encouragement of Dr. Ophélie Ronce has been invaluable.

Thanks to the zoology department and the members of the Biodiversity Research Center, for creating an open and welcoming workplace. I would not have come far with my bioinformatics chapter without the help from the zoology computing unit, in particular Dr. Alistair Blachford.

I'm grateful to my funding sources: the Four-Year Scholarship from UBC, funding from the National Science and Engineering Research Council of Canada (Discovery grant to S.P. Otto), and St Johns College in the form of Charles C.C. & Sophia Wong Memorial Fellowship and George Shen Fellowship.

Lastly, a big thank you to my friends and family, who supported me along the way. In particular, Brooke Xiang, Szu Shen, Dr. Ailene MacPherson, Hanna Carlsson-Murray, Matthew Cravens, Edgar Liao, Barbara Neto-Bradley, Léonard Dekens, Mackenzie Urquhart-Cronish, Ken Thompson, and Isabelle Berg. My parents, my grandparents, my siblings, and my niblings, who make me smile. All my love to Gil Henriques, who carried me through the rough patches.

Dedication

To Newton,
who appeared in my life when I needed it the most,
and took the leading role in my dreams

Chapter 1: Introduction

Life is mysterious and yet, the theory of evolution through the means of natural selection explains the vast diversity of life forms on earth. To test our intuition for the conditions that would favor one trait over another, we can experiment with organisms that have short life spans, permitting us to observe evolution over many hundreds or even thousands of generations.

The power of experimental evolution has been used to measure benefits and costs of sexual reproduction. Two unicellular, facultatively sexual organisms have been used extensively in this research: the budding yeast *Saccharomyces cerevisiae* and the green algae *Chlamydomonas reinhardtii*. Even though they come from different kingdoms they share features that make them especially useful for evolutionary experiments regarding sexual reproduction. Firstly, they can be propagated both clonally through mitotic replication and sexually through meiosis. This makes it possible to evolve populations under different reproductive modes in the lab and compare the evolutionary outcome, which has been done both in *C. reinhardtii* (Kaltz and Bell 2002; Renaut et al. 2006; Lachapelle and Bell 2012) and *S. cerevisiae* (Zeyl and Bell 1997; Gray and Goddard 2012; McDonald et al. 2016). Secondly, both *C. reinhardtii* and *S. cerevisiae* exist as two different mating types in their haploid state. In *C. reinhardtii* these are referred to as mt⁺ and mt⁻, while in yeast they are called MAT^a and MAT^α. Cells of the different mating types are indistinguishable under the microscope (they are what is called isogametes) but possess different pheromones and receptors such that mating is possible only between cells of opposite mating types (for a review in *Chlamydomonas* see Goodenough et al. 2007, for *Saccharomyces* see Haber 2012). The existence of mating types in these unicellular eukaryotes makes them suitable to address questions regarding the evolution of different sexes and the evolution of differentiated gametic forms for the two sexes (anisogamy) from an isogametic ancestor (Charlesworth 1994). The second chapter of my thesis

explores both the evolution of reproductive modes and mating competition by studying responses to different mating ratios and to an increased frequency of sexual reproduction, and I give a brief overview of previous work in this area in section 1.1.

Even though many experimental evolution studies concentrate on adaptation to novel environments by enforcing novel selective regimes, both *S. cerevisiae* and *C. reinhardtii* have also been used in evolutionary experiments where the aim is to study the creative process of evolution, namely mutations, in the absence of selection. By maintaining populations at low density (to prevent competition between cells) and randomly selecting individuals to produce the next generation, mutations that would usually be eliminated by selection will accumulate in the population, an experiment called mutation accumulation (MA). Not only do these experiments tell us of the rate and kind of mutations (such as single nucleotide or insertion/deletion of multiple nucleotides) arising in the genomes of *S. cerevisiae* (Lynch et al. 2008; Zhu et al. 2014; Sharp et al. 2018; Liu and Zhang 2019) and *C. reinhardtii* (Ness et al. 2012, 2015; Sung et al. 2012; Morgan et al. 2014), they can also be used to explore the fitness effects of spontaneous mutations (Kraemer et al. 2017; Böndel et al. 2019). In the third chapter of my thesis, I explore if we can use a bioinformatic tool that predicts functional effects of mutations on proteins, PROVEAN, to explain fitness differences among MA lines in three different datasets: two from MA in *S. cerevisiae* (Sharp et al. 2018; Liu and Zhang 2019) and one from MA in *C. reinhardtii* (Morgan et al. 2014; Ness et al. 2015). I describe how others have approached the challenge of linking genotype and protein function to fitness in section 1.2.

Describing mutations that arise in the absence of selection informs us of the raw material involved with evolution. When we have an expectation for the kind of mutations that should arise and fix in populations at random, we will be able to tell how selection has acted in populations by

searching for outliers to our expectations. One of the most fundamental questions about the phenotypic and fitness effect of mutations regards the extent to which a mutation affects fitness when present in a diploid genome. As Mendel was able to tell from breeding peas of different colors, some mutations have the same effect whether they are present in a single (heterozygous) or double copy (homozygous) in the organism. These mutations were named dominant, in contrast to mutations whose effects on flower color were seen only when present in double copy: these were referred to as recessive. Noteworthy, Mendel found that the wildtype phenotype was dominant to the mutant phenotype in many cases. I describe the reigning explanation for the dominance of wildtype alleles in section 1.3. Mutations can have different effects on fitness when first arising in a population (when individuals will be heterozygous for the mutation) than when established (when the majority of individuals will be homozygous), the combination of the two determines the likelihood that mutations will rise and fix in populations. Knowing the effect of mutations in only one of these states (heterozygote or homozygote) is therefore insufficient to construct an accurate expectation of the fitness effects of mutations that accumulate in natural populations. Likewise, mutations in sexually reproducing organisms may have different effects on fitness when in the haploid (single copy) state and the diploid homozygous state. I suggest reasons for why we would expect such differences and empirical evidence supporting it in section 1.3.

1.1 Experiments on sexual selection

Sexual reproduction is defined by the reduction in chromosome copy number through meiosis and subsequent union of haploids. While our view of sexual selection is dominated by the extensive differentiation in both morphology and behavior of the sexes in multicellular organisms, unicellular fungi can elucidate the emergence of more complex forms of sexual differentiation. Most sexual unicellular eukaryotes have two or more different mating types (Hadjivasiliou and

Pomiankowski 2016). While in some genera the cell types are of similar size and shape (isogamy), in others they are differentiated in either motility and/or morphology (anisogamy). The most well-accepted hypothesis for this diversity (the PBS theory, Parker et al. 1972) states three necessary conditions for the evolution of anisogamy: disruptive selection, limited reproductive resources, and a relationship between zygote size and fitness. Anisogamy will evolve when a reduction in the size of the male gamete increases the likelihood of fusion at the same time as an increase in the size of the female gamete increases zygote survival (Lessells et al. 2009). Depending on the environment and on how the fitness of the zygote relates to its size this can entail a decrease in size in favor of being numerous for all mating types or differentiation into a sessile and motile form (Bulmer & Parker, 2002).

Sexual selection theory is founded on the assumption of differentiated gametes. The ratio of male to female gametes determines the strength of competition for mating opportunities. The less common sex will be more guaranteed to mate and can be picky, whereas the more common sex will have to compete for mating opportunities. However, competition for mating opportunities can exist and determine evolution even in isogamous species. In wild *Saccharomyces*, haploids of the two mating types are produced as spores in close association with its siblings, and they often mate/fuse with the cell in their immediate proximity (Knop, 2006, though see McClure, Jacobs, Zyla, & Lew, 2018). However, if a haploid yeast cell finds itself surrounded by cells of the same mating type, it also has the ability to switch mating type (homothallism) and mate with its prior daughter cell (Haber 2012). The fact that yeast has evolved this function is telling of the strength of selection for mating. Strains of yeast that had lost this ability to mate-type switch (heterothallism) were selected to maintain many laboratory strains. These can therefore be kept as haploids of a single mating type and allow for experiments of mating competition by varying the

ratio of one mating type to another. Such experiments provide insight into the evolution of differentiated gametes under asymmetric mating opportunities.

Experimental studies of sexual selection in *S. cerevisiae* have provided important insights about the evolution of reproductive diversity. Rogers and Greig (2009) performed an evolution experiment with two different kinds of *MAT α* cells: they either had a normal or weakened production of pheromone that attracts the opposite mating type, the *MAT α* cells. The different types of *MAT α* cells were initially present in all experimental cultures, and the purpose of the experiment was to test if the strong signaler will always evolve to fix in the population. For 13 consecutive sexual cycles they mated their experimental cultures either with an equal ratio of *MAT α* to *MAT α* cells, or with a surplus of *MAT α* cells. The abundance of *MAT α* cells in the latter treatment would lead to stronger competition for mating opportunity. In accordance with sexual selection theory, they found that the strong *MAT α* signaler rose in frequency and fixed in the experimental cultures that were mated with a biased ratio of the two mating types, while it failed to do so when the mating types were mated in equal ratio. Hence, this study by Rogers and Greig (2009) proved that budding yeast could be used to test theories of sexual selection.

In another experiment, Smith et al. (2013) showed that haploid spores of yeast have mating preferences even within a mating type. The experimenters first determined the relationship between spore size and fitness in rich and poor growth medium. They found that larger-sized spores (regardless of mating type) germinated quicker on rich medium and that smaller-sized spores germinated quicker on poor medium. They also showed that the diploid resulting from mating of two large-sized spores divided more rapidly on rich medium than did a diploid resulting from the mating of two small-sized spores, and vice versa. They then set up mating trials where a focal cell (either large or small) was placed with a large and a small-sized spore on either rich or

poor medium and observed which of the spores the focal spore mated with. Interestingly, the focal cell (regardless of whether it was a small-sized or large-sized spore) preferentially mated with the spore who was of the size that was favorable in the current environment. The result from the Smith et al. (2013) study showed convincingly that yeast cells are able to choose their mating partner.

In the second chapter, I subjected populations of budding yeast, which primarily reproduces clonally, to a forced change in life history by increasing the frequency of sex. We varied the extent of sexual competition during mating by combining either equal or biased ratios of the two mating types. The populations evolved increased mating rate and asexual growth and decreased haploid cell size. In treatments with strong sexual competition, the common mating type showed a trade-off between increased asexual growth and increased mating rate, suggesting that increased “gamete” numbers and increased attractiveness arose as alternative strategies in response to intense sexual competition. While the two mating types of yeast are often described as indistinguishable, crosses between haploids from independent populations suggest that one mating type responds more strongly to sexual selection, which I discuss further in Chapter 2.

1.2 Linking genotype to fitness

Many population geneticists search for loci underlying adaptation by comparing divergent populations. Locations in the genome with high differentiation between populations and low diversity within populations, so called “selective sweeps”, are often taken as candidate loci for adaptation (Stephan 2019). There are also sections within the genome that seem to be conserved across species, genera, and taxa (Siepel et al. 2005). These sections signal strong constraints to the function of the sequence (typically transcribed genes). Even within these genes, some sites are more constrained than others. Henikoff & Henikoff (1991) assembled conserved sequences of proteins and computed the likelihood of any amino acid substitution in these sequence blocks

(called the BLOSUM matrix). The BLOSUM matrix has since been used to score homology between proteins. To know how closely related two sequences are, one can align them and score any mismatching amino acids based on their likelihood of substitution from the BLOSUM matrix. While it is intuitive that purifying selection rids core house-keeping genes of new mutations (as has been found across taxa, Harris et al. 2003), it is less evident that it will act in a similarly constrained way in other proteins. Ng & Henikoff (2001) argued that it is possible to infer the functional effect of mutations in a protein based on the frequency of that change in similar proteins. Their method and others that expand upon it (Choi et al. 2012) are widely used to find medically relevant protein variants in humans. Within the past five years or so, this approach has also been gaining traction in evolutionary biology. If protein comparisons can predict disease-causing alleles in humans, why not apply it to natural populations of other species? Lind et al. (2017) tested the correlation between the prediction of a particular tool using sequence alignment, PROVEAN, and the growth rate of single-gene mutants in the bacteria *Salmonella typhimurium* and found mixed results. The functional prediction of PROVEAN correlated with fitness differences among mutants of a particular set of genes, but not others. It is not evident how the predictions for functional impacts of mutations on proteins across the genome can be used to estimate fitness differences between populations.

In the third chapter, we use previously published data of MA lines with known mutations and growth rates to assess whether PROVEAN has the power to predict fitness effects of mutations accumulated in unicellular laboratory organisms, two in *S. cerevisiae* (Sharp et al. 2018; Liu and Zhang 2019) and one in *C. reinhardtii* (Morgan et al. 2014; Ness et al. 2015; Kraemer et al. 2017). We find that mutational effects based on PROVEAN score have little to no predictive power in the three datasets examined, which involve different ploidy levels, environmental conditions, and

genetic backgrounds. Furthermore, we find that the default parameter choices of PROVEAN may be biased toward species that are better represented in sequence databases. We believe that the reason behind this low predictability could be because there is no accounting of the importance of each protein to the fitness of the organism.

1.3 Mutational effects matter

Mendel's (1865) work is significant not only because he found patterns of particulate inheritance, founding our understanding of genetics, but also because he demonstrated that mutations most often were recessive. Rather than the 1:2:1 ratio of wildtype, intermediate, and mutant phenotypes, he often observed a 3:1 ratio with dominance of the wild type allele. While Fisher (Fisher 1928, 1931) originally suggested the recessiveness of mutants to be the result of evolutionary adaptation, Wright (1934) showed that the pattern could be explained by redundancy in metabolic pathways: the production of wildtype protein from one functional copy of the gene would be sufficient to produce a wildtype phenotype. This idea was later developed by Kacser and Burns (1981) and termed the metabolic control theory of dominance. In short, they imagine a pathway of enzymes, where the product of one enzyme becomes the substrate of the next. If we assume that the concentration of substrate is low relative to the concentration of enzyme we can describe the rate of production of the end product, which is referred to as flux, in terms of a set of linear differential equations. These describe the rate of conversion at each step in the metabolic pathway. Solving these equations for the case where the flux is constant, at steady state, shows that the dependence of the flux on any particular enzyme hinge on the number of enzymes in the pathway (Figure 1.1A). The more enzymes a pathway consists of, the less sensitive the flux will be to a reduction in the concentration or efficiency of one of the enzymes (with the exception of the most limiting enzyme with the highest “control coefficient” over flux). Kacser and Burns (1981) describe that

because most enzymes are part of pathways with more than one reaction step most mutations will not lead to a stark difference in phenotype, hence explaining the dominance of wildtype phenotypes. Only in cases of severe mutations such as a null mutant that sharply reduces the enzyme concentration or function would we expect the mutant phenotype to be detectable Figure 1.1B.

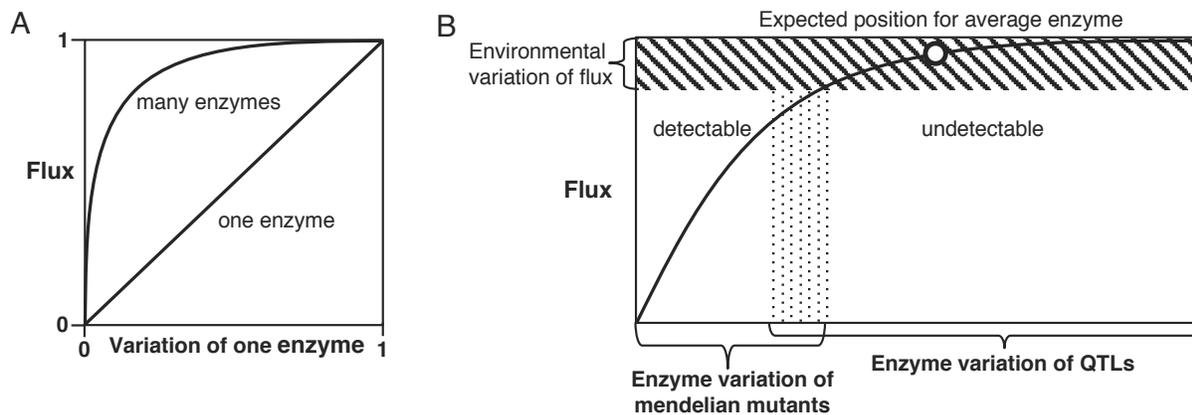


Figure 1.1 Dominance of wildtype phenotype is an inherent property of multi-step enzyme pathways. A: In one step enzyme pathways, the variation in enzyme concentration or function will correspond linearly with the rate of production of the end product (flux). When a pathway consists of many enzymes, a change in the concentration or function of one of the enzymes in that chain will have a hyperbolic relationship to the flux. B: Only mutations that seriously affect the concentration or functionality of an enzyme will be detectable at the level of the phenotype (flux). Variation in quantitative traits can be explained through polymorphisms affecting protein function and concentration that are undetectable on the level of the phenotype. Reproduced from Kacser and Burns (1981).

The dominance of fitness effects of spontaneous mutations plays an important role in our understanding of evolution. It can determine the likelihood of evolutionary rescue (Unckless and Orr 2020), the fixation probability of new mutations (Morton 1971; Whitlock 2002), the evolution of ploidy levels (Orr and Otto 1994; Otto 1994; Scott and Rescan 2017), and the evolution of sex (Chasnov 2000; Otto 2003). There have been several empirical estimates for the dominance effects of deleterious mutations (Agrawal and Whitlock 2011; Manna et al. 2012; Marek and Korona

2016). The limit to these studies is that they have focused on large-effect mutations (such as gene knockouts), which are typically rare.

Separately from the question of dominance is the question of ploidy-specific effects of mutations. In a similar way to how dimorphism can evolve between sexes of the same species, the haploid and diploid form of a species can look identical (isomorphic, for example alga such as *Ulva lactuca*) or different (heteromorphic, for example ferns such as *Onoclea sensibilis*). The few models that consider the effect of selection in both the haploid and diploid form of the life cycle often assume that the fitness effect of a mutation is identical in the haploid and homozygous diploid individual (Mable and Otto 1998). Making this assumption allows researchers to estimate an “effective dominance” by comparing the fitness in heterozygotes ($1-hs$) to that in haploids ($1-s$). Dominance has been shown to determine the evolution of haploid-dominant or diploid-dominant life cycles (Scott and Rescan 2017). For example, when measuring the dominance effect of mutations, Marek and Korona (2016) used haploids rather than homozygous diploids to measure the selection coefficient, clearly stating that the two can be assumed to be close to identical. This is counter to the few empirical tests that have been conducted which show that mutational effects may differ between ploidy levels (Mable 2001; Zörgö et al. 2013).

In the fourth chapter, I attempt to measure the fitness impacts of spontaneous small-effect mutations in haploid and diploid yeast, without assuming that haploids and homozygote diploids are equivalent, again using the MA lines of Sharp et al. (2018). We found that the method used to transform the yeast into the three genotypic states introduced mutations with large effects on fitness that overwhelm the effects we attempted to measure. While these results were disappointing, we believe that they are important, because plasmid transformations of the kind we conducted in this chapter are standard practices in most yeast labs. We believe that the issue of *de*

novo mutations during transformations and mating that occurred in our experiment would likely affect other experimental evolution studies attempting to compare fitness via genetic manipulation of strains.

Taken together, my work demonstrates the utility of yeast to ask questions about the evolution of life cycles, sexual reproduction, and the limit of both experimental and *in silico* methods to characterize the path from genotype to fitness.

Chapter 2: Evolutionary responses to increased sexual selection in yeast

2.1 Introduction

Species adapt to their reproductive strategy. Organisms that engage in outcrossing sex need to find a mate and go through meiosis. In obligately sexual species, mutations that increase an organism's ability to perform in this arena are readily visible to selection. Sexual selection is responsible for much of nature's diversity: from the hypersensitivity of moth antennae for the pheromone of the opposite sex, to the stunning precision with which corals coordinate the release of gametes. In facultatively sexual species, sexual traits can be hidden from selection for many generations. Whatever traits and alleles that are favored in one bout of sex may be lost through the period of asexual reproduction that follows.

The evolutionary benefit of specialization is evident from the divergence into obligate asexual and sexual species from an ancestral facultatively sexual species (Schurko et al. 2009). It can also be seen when comparing laboratory populations of budding yeast to their wild counterparts: domesticated strains of *Saccharomyces cerevisiae* typically have much lower success in initiating sporulation (meiosis) compared to wild isolates (Gerke et al. 2006). The loss of rarely used traits can be either because of a lack of constraint (new mutations that reduce the trait are selectively neutral) or because of selective trade-offs (either due to pleiotropic or energetic dependence between traits). There are bountiful examples of the second proposal. Similar to the well-known trade-off between an animal's size and their rate of reproduction, the energy that facultatively sexual organisms would spend on finding a mate, gamete fusion, and meiosis could be otherwise spent on asexual growth. An increased intensity of sexual selection could therefore result in a reduction in non-sexual traits, including rates of mitotic division and cell size.

Experimental evolution gives us an opportunity to determine the evolutionary consequences of a change in mating system, as a way of identifying key traits and trade-offs.

2.1.1 Yeast as a system to study sexual selection

The properties of *S. cerevisiae* have made it a popular system for evolutionary studies of reproductive mode. Natural isolates of yeast are most often diploid (Peter et al. 2018). In response to nutrient starvation, diploid cells undergo meiosis and sporulation to produce haploid offspring, which bear either *MATa* or *MAT α* mating types. The two mating types look identical but produce different pheromones, called a- and α -factor, respectively. They also possess pheromone receptors for the factor produced by the opposite mating type. After sporulation, the haploid cells germinate and start growing once favorable conditions return. Haploid yeast can propagate asexually but arrest growth and mate when in the presence of the opposite mating type. Because each diploid meiosis results in two spores of each mating type, it is believed to be common for siblings to mate with one another immediately upon germination, referred to as automixis (Knop 2006). Outcrossing sex (mating between lineages that are not immediate relatives) is estimated to have occurred only once every 50 000 asexual generations in this species (Ruderfer et al. 2006; Magwene et al. 2011), implying that, as a fraction of total selection, sexual selection is probably weak. We can prevent immediate mating by separating the two mating types before germination and alter the strength of sexual selection by controlling the time allowed for mating as well as the ratio of the two mating types.

There is evidence that trade-offs shape patterns of sexual traits in *S. cerevisiae*. Deleting the genes that produce pheromones (*mfa1*, *mfa2*, *mfa1*, and *mfa2*) increases growth rate in both mating types (Smith and Greig 2010). Lang et al. (2009) found that sterile yeast have increased asexual growth, probably due to a decrease in costly expression of genes not directly involved in

the mating pathway (Lang et al. 2009). In contrast, deleting genes responsible for meiosis gives no direct benefit to asexual growth (Goddard et al. 2005). It thus seems plausible that the loss of meiotic capacity in domesticated and asexually propagated strains is not due to pleiotropic mutations that increase growth rate at the expense of sporulation success, but rather because sporulation-reducing mutations are effectively neutral under asexual culturing conditions. Zeyl et al. (2005) observed reduced sporulation rates in laboratory populations subject to prolonged asexual propagation but were able to recover sporulation rates in some populations through repeated cycles of sexual propagation.

Apart from how often a species engages in sex, the way in which mating occurs also matters for the evolution of sexual traits. The trade-off between sexual and asexual traits in yeast will be resolved differently depending on the mating system. Reding et al. (2013) found that populations experiencing weak mating competition adapted faster to a new environment compared to populations that underwent strong sexual selection. Likewise, impaired mating capability in yeast can be recovered only if there is strong competition for mates (Rogers and Greig 2009).

Our aim was to measure changes in growth rate, mating rate, sporulation rate, and cell size following evolution under different strengths of sexual selection in initially isogenic populations of yeast. Manipulating the ratio of the two mating types alters the intensity of sexual selection. The common mating type that experiences stronger selection can respond through increased mate attraction or by producing more “gametes” than competing lineages. An increase in the rate of cell division could result in a reduction in cell size. Previous studies of mating system in yeast have either had equal mating-type ratios (Zeyl et al. 2005b) or enforced strong sexual selection for only one of the two mating types (Rogers and Greig 2009; Reding et al. 2013). As far as we are aware, no previous study has evolved populations under different mating-type ratios and estimated the

response in mating rate. In addition, we also measure how non-sexual traits (growth rate and cell size) respond to increased sexual selection.

2.1.2 Can yeast adapt to their mating system?

To understand how populations adapt to their mating systems, we evolved budding yeast under different mating conditions (Figure 2.1). We used the experimental system previously developed by McDonald et al. (2016) to force replicate populations to go through frequent rounds of sex with controlled periods of haploid growth, mating, diploid growth, and sporulation. During mating, *MAT_a* and *MAT_α* cells were combined either in equal ratios, with a surplus of *MAT_a* (a-biased), or with a surplus of *MAT_α* (α-biased). We also included a treatment where the mating types were not isolated during germination but rather were allowed to immediately mate. This treatment imposes weak sexual selection given the prolonged contact of *MAT_a* and *MAT_α*. We refer to this control as the immediate mating (IM) treatment, as distinguished from the mating ratio treatments, which we refer to as delayed mating (DM).

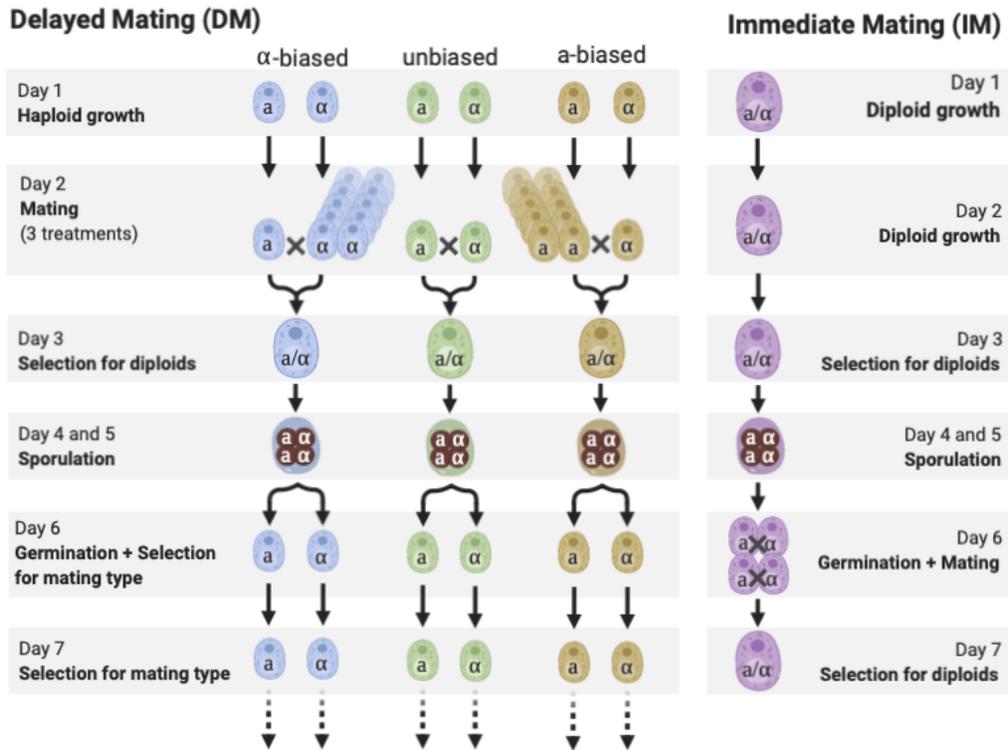
We predicted that yeast would become faster at mating to different extents in the different treatments. Populations in the skewed mating ratios were expected to obtain mutations favoring competitive mating ability in the common mating type cells, *e.g.* by increasing production of mating pheromone or sensitivity to the opposite mating factor. We also investigated whether both mating types of a population evolved to be better at mating in general, regardless of the evolutionary history of their partner, or whether the mating types coevolved to the particular mating condition.

After ten rounds of sexual reproduction, we measured the mating rate of genotypes sampled from each population under the three different mating ratios to test whether they had adapted to their mating treatment. We found that all lines had increased their mating rate compared to the

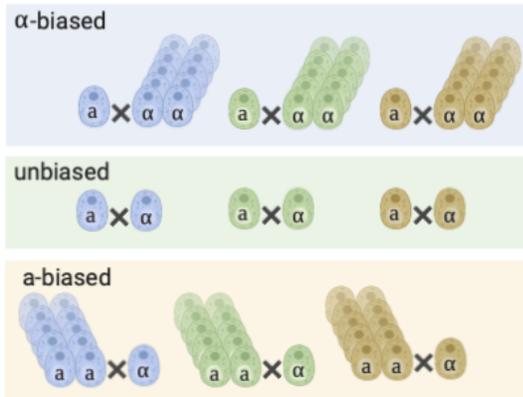
ancestor when mated at an equal ratio, but large differences in mating rate between evolutionary treatment groups in the biased ratios. We discuss the differences in the mating machinery of *MATa* and *MATα* cells that could drive this pattern. We also measured changes in asexual growth rate in both the diploid and haploid versions of our lines, as well as changes to cell size in both mating types. We found no evolutionary change in sporulation rate, despite strong and frequent selection on this trait.

2.2 Methods

A. Evolutionary experiment



B. Mating rate assay



C. Cross-couple mating rate assay

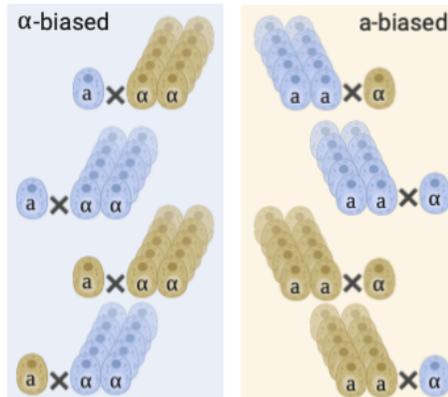


Figure 2.1 Schematic of evolutionary experiment (A) and post-evolution mating rate assays (B and C). Appreciable growth occurred on all days apart from days 4-6 in A. Assay B tests for adaptation to the mating environment. Assay C tests for mating type-specific adaptation, as well as co-evolution of the two mating types. The background color in B and C represent the ratio of $MATa$ and $MAT\alpha$ in the mating culture (blue is excess of $MAT\alpha$, green is unbiased, and yellow is excess of $MATa$). The color of the cells signifies the evolutionary history of the line used. Created with BioRender.com.

2.2.1 Strains

We used the strain MJM64, with genotype *MATa YCR043C::KanMX*, *ho*, *Pr_{STE5}-URA3*, *ade2-1*, *his3Δ::3xHA*, *leu2Δ::3xHA*, *trp1-1*, *can1::Pr_{STE2}-HIS3-Pr_{STE3}-LEU2* and strain MJM36 with genotype *MATα YCR043C::HphMX*, *ho*, *Pr_{STE5}-URA3*, *ade2-1*, *his3Δ::3xHA*, *leu2Δ::3xHA*, *trp1-1*, *can1::Pr_{STE2}-HIS3-Pr_{STE3}-LEU2* described in McDonald et al. (2016). Because the *MATα* strain received had a petite phenotype, we crossed the *MATa* and *MATα* and sporulated the resulting diploid to restore functioning mitochondria. Through tetrad dissection and subsequent detection of auxotrophic and antibiotic markers, we selected the segregated products with the desired parental genotypes. The strains were designed to allow for selection of specific mating types and ploidy levels, using both auxotrophic and antibiotic markers (McDonald et al. 2016 and A.1).

2.2.2 Media

We combined selection on the auxotrophic markers and the antibiotics to select for haploid cells. Because ammonium sulphate (AS), which is commonly used as a nitrogen source in synthetic medium, interferes with the antibiotics we use, we replaced it with monosodium glutamate (MSG) in our media (referred to as SE). The media used in the experiment can be found in Table 2.1.

Table 2.1 Media used in this experiment

Medium	Selects for
SE ^a -Ura -His + g418 ^b	Haploid <i>MATa</i>
SE ^a -Ura -Leu + hygro ^c	Haploid <i>MATα</i>
SE ^a + 5-FOA + g418 ^b + hygro ^c	Diploid <i>MATa</i> / <i>MATα</i>
SE ^a + 1 % potassium acetate	Presporulation
Essential amino acids ^d + 1 % potassium acetate	Sporulation

^a Dextrose (42 g L-1) + YNB (3.9 g L-1) + MSG (3 g L-1)

^b 250 μg mL-1

^c 200 μg mL-1

^d 5 mg L-1 of Ade, Ura, His, and Trp. 25 mg L-1 of Leu.

2.2.3 Evolution experiment

All growth took place in 1 mL liquid cultures in 96 well plates at 30 °C with shaking. We founded 96 *MATa* and 96 *MATα* populations by inoculating 10 μL of saturated culture of each ancestor in 1 mL of mating type-specific medium. We compare results between the DM and IM lines and among the DM lines, all with reference to the properties of their ancestors. We did not include an asexual control because of the need to observe growth in two types of haploid-selective media and diploid-selective medium and because we were concerned about ploidy changes, as observed in previous experiments (Gerstein et al. 2006; Gerstein and Berman 2015).

The populations were put through ten sexual cycles, each of which took 7 days to complete (Figure 2.1). Samples were stored and later frozen during each cycle at day two before mating (haploids) and day 3 before transfer to presporulation medium (diploids). After the seventh day of the tenth sexual cycle, populations were stored and frozen. There was appreciable growth during four days per sexual cycle. Given a doubling time of 90 minutes under optimal conditions, we expect a maximum of 64 asexual generations per experimental cycle.

2.2.3.1 Mating

The cultures of opposite mating types were mated in one 96 well plate. Mating type ratio was manipulated by altering the relative volumes of *MATa* to *MATα* culture used: either combined in an equal ratio or 1:10 ratio in favor of either *MATa* or *MATα*, with 24 replicate populations for each of these three treatments. We chose to hold constant the total number of potential diploids, which resulted in a weaker dilution for the more common mating type in the skewed mating ratio treatment (see Table 2.2).

Table 2.2 Mating ratios during the experiment and in post-experiment assays

Ratio	<i>MATa</i> [mL] ^a	<i>MATα</i> [mL] ^a	<i>MATa</i> dil ^b	<i>MATα</i> dil ^b	Cell dil ^b
a-biased	1	0.1	5.5	55	5
unbiased	0.1	0.1	50	50	25
α-biased	0.1	1	55	5.5	5

a Volumes sampled from the saturated cultures of haploids grown in SE.

b Final dilutions of each mating type or of the total number of cells (see details on dilution series in Appendix A.2)

To match the DM treatments, the 24 populations of the IM treatment were grown in two different sets of wells during the periods where the DM lines were grown as separate mating types and then combined (in equal ratio of the two sets, similar to the equal mating ratio treatment).

2.2.3.2 Sporulation

After 15 h of growth in presporulation medium (5 pm – 8 am), the cultures were all transferred to individual 10 mL plastic tubes, where they were spun down, supernatant removed, and 1 mL of sporulation medium was added. The tubes for sporulation were left in a roller drum at room temperature for two nights (48 h).

2.2.3.3 Ascus digestion and germination

The sporulated cultures were all transferred to microcentrifuge tubes, spun down, and supernatant removed. We treated the sporulated lines with 5 μL of 5 $\text{U } \mu\text{L}^{-1}$ of zymolyase with 45 μL of sterile water (giving 0.5 $\text{U } \mu\text{L}^{-1}$) for 20 minutes at 35 °C, after which the tubes were put on ice. The DM cultures were sonicated on high on a Bioruptor machine with 30 seconds of sonication followed by 30 seconds of rest, repeated ten times. The sonicated cultures were then divided, with 20 μL used to inoculate 1 mL of *MATa*-specific media and 20 μL used to inoculate 1 mL of *MATα*-specific media. For the IM treatment, after the zymolyase digestion, we did not sonicate the cultures as we wanted the tetrads to remain intact to allow for within-tetrad mating. Rather, we split the culture into two wells, transferring 20 μL culture into 1 mL SE medium.

After two days of germination and growth in mating type-specific media for the DM lines or one day of germination and growth plus one day of growth in diploid-specific medium for the IM lines (days 6 and 7 in Figure 2.1A), all lines were diluted 1:100 into SE medium, and the cycle restarted.

2.2.4 Assays

At the end of the experiment, the frozen cultures from day 3 (diploids) and 7 (haploids) of cycle 10 were streaked out on SE plates. Single colonies were inoculated into 2 mL of SE medium to obtain isogenic representatives for each population. After 2 days of growth, cultures were combined with glycerol and frozen at -80 °C. This procedure was also done to eight independently grown cultures of the ancestral haploid *MAT_a*, the ancestral haploid *MAT_α*, and a diploid control strain formed by mating the two ancestors. The cultures of this diploid strain were only used as a control in flow cytometry when assaying genome and cell size of the haploid evolved lines. The cultures of haploid ancestors were also used in flow cytometry and in the other assays of the evolved haploid lines. The diploid lines of cycle 1 (day 3) were independently grown up from single colonies and frozen and are used as the ancestral control when assaying the evolved diploid.

2.2.4.1 Mating rate assays

The mating rates of lines were measured under unbiased, α -biased and a -biased ratios (Table 2.2) with three replicates each. Haploids were grown in SE medium overnight before being combined. After 24 h, mating cultures were diluted and plated on permissive and diploid-selective media. The assay was conducted in 9 blocks. We used SE and SE + 5-FOA + g418 + hygromycin for permissive and diploid-selective media in the first two blocks, respectively, but switched to YPAD (Yeast Peptone Dextrose with extra 40 mg L⁻¹ of Adenine added) and YPAD + 5-FOA + g418 + hygromycin in the remaining blocks, for practical reasons (block was used as a random effect in

our model). Because only a small fraction of cells mated, we made separate dilutions for plating on permissive and drug plates. After two days of growth, colonies were counted.

To calculate mating rate we first converted colony counts into cell concentrations in the mated culture. The selective plates gives the concentration of diploid cells, D_{mated} , whereas the permissive plates gives the total concentration of cells, C_{mated} (a mix of haploids and diploids). The maximum concentration of diploids, D_{max} , is different for the biased and unbiased mating. In unbiased mating ratio, the final culture could (theoretically) consist of 100% diploids ($D_{\text{max}} = C_{\text{mated}}$). In contrast, due to the excess of one mating type, the maximum fraction of diploids in the biased ratio would be 10% because 10:1 mating ratio would produce at most 9:1 haploids:diploids if all the rare types mated ($D_{\text{max}} = 0.1 C_{\text{mated}}$). Mating rate was calculated by dividing D_{mated} by D_{max} . This calculation assumes that all cell types grew at the same rate during the 24h mating period. Because the presence of *MATa* and *MATα* pheromones inhibits haploid growth, diploid cells may have grown faster once produced, which would bias the estimated mating rate upwards.

To avoid cases where we undersampled the diploids and therefore have a noisy estimate of mating rate, we calculated the maximum number of diploids that we could have plated on the selective plate had there been 100% mating (D_{max} /the dilution factor used for the selective plate) and removed any estimates where this number was below 500.

Because of the way we infer mating rate (the concentration of diploids from one plate and the maximum possible concentration of diploids from another), the error distribution is not binomial (which would normally be used for sampled proportions). We thus fit a linear-mixed effect model to our data, using line ID and block as random effects and evolutionary treatment and assay mating ratio as fixed effects, allowing interactions. To disentangle the interactive effects of

evolutionary treatment and assay ratio we use the emmeans package (Lenth 2020) to make conditional pairwise comparisons between assay ratios and treatment groups.

We tested whether changes in mating rate were due to evolutionary changes in the common or in the rare mating type by doing cross-couple matings. We considered only the skewed mating ratio treatments. From each treatment group, we chose three populations exhibiting higher than average mating rates in the mating ratio to which they had evolved. The six *MATa* lines and *MAT α* lines were mated in a total of 36 combinations. Mating rate was measured three times, split over two blocks, in the two different mating ratios (a-biased or α -biased). Because we had observed a higher mating rate in the α -biased ratio in our previous assays we diluted these mating cultures 10-fold more than the cultures mated in the a-biased ratio. Because the mating rate was very low, we decided to exclude measurements from plates where the maximum number of diploids that could have been plated on the selective plates (D_{\max} /the dilution factor used for the selective plate) was below 50. To test whether the increase in mating rate of the couples was achieved by adaptation to high competition of the more common mating type (*MATa* lines in the a-biased treatment and *MAT α* lines in the α -biased treatment), we fit a linear mixed-effect model with mating couple as a random effect and with assay mating ratio, the treatment under which the *MATa* line evolved, and the treatment under which the *MAT α* line evolved as fixed effects. We allowed for three-way interactions in this model. We assess the effect of the interactions with pairwise comparisons (conditional on one of the three factors) using the emmeans package (Lenth 2020).

To estimate the mating rate in the IM treatment, we sporulated the ancestral diploid in 16 replicates, treated the cultures with zymolyase with the same procedure as we used for the IM treatment during the evolution experiment, and germinated them in the same ratio of treated culture to medium as we used during evolution. After 24 hours of growth we diluted the culture and spread

it on permissive and diploid-selective plates. Using colony counts on the two types of plates we calculated the mating rate as described above.

2.2.4.2 Sporulation rate assays

Sporulation rate was measured for 32 lines during the experiment (8 lines from the IM treatment, 8 from each of the DM treatments). For each sexual cycle, we sampled 10 μ L of sporulated culture and counted the number of cells and the number of tetrads with the use of a hemocytometer (tetrads were counted as one cell). Around 150 cells were counted per line per assay (mean = 150.2, standard deviation = 52). Sporulation rate was calculated as the number of tetrads divided by the total number of cells counted. We evaluated a generalized linear model of sporulation rate explained by evolutionary treatment and the number of sexual cycles and allowed interactions between treatment and the number of sexual cycles.

After the end of the experiment, we grew the diploid lines from the DM treatments of both sexual cycle 1 and sexual cycle 10 freezer stocks. The cultures were sporulated for two days using the protocol described above. The assay was conducted in six blocks and spores were counted by two people. Around 300 cells were counted per line (mean = 302, standard deviation = 127). Because block and counter ID had low variance associated with them and because we encountered convergence errors when attempting to include these in our model, we excluded these random factors and instead fit a generalized linear model to sporulation rate as a function of evolutionary treatment or ancestor.

2.2.4.3 Growth-rate assays

All lines were grown from frozen stocks by adding 20 μ L of stock to 1 mL of SE medium, kept in shaking 96-well boxes for 2 days at 30 °C. Lines were then transferred to individual microcentrifuge tubes and held at 4 °C to arrest growth. For every growth-rate assay, the lines

were diluted 1:121 and run on a Bioscreen C for between 19-24 h at 30 °C with continuous shaking, taking an OD₆₀₀ reading every 15 minutes.

We estimated growth rate as the maximum slope of a spline fit to log-transformed OD readings using the loess function in R (Gerstein 2012). We tested differences between the evolved lines of different treatments using ANOVA and compared evolved to ancestral lines with Welch's t-test.

2.2.4.4 Flow cytometry

We used flow cytometry to assess cell size and genome size. The haploid lines and their ancestors (together with a diploid control) were prepared for flow cytometry by fixing with ethanol and staining with Sytox Green using a protocol similar to Gerstein et al. (2006) (A.3). Samples were run on an Attune NxT analyser, using 200 µL of the stained cells in 96-well plates. Running on the plate setting, 160 µL of each sample was processed at 200 µL per minute.

The FCS files were exported and analyzed with the flowCore package (Ellis et al. 2019) in R. After removing events with very low and high fluorescence or forward scatter, the location of the first fluorescence peak (BL1.H) was taken as the genomic content of cells. We then took the standard deviation in BL1.H of all events and selected cells with fluorescence within the range $\text{peak1} \pm 0.5 \text{ sd}$. We used the average forward scatter (FSC.H) of these events as a measure of cell size. Two lines were excluded from the analysis due to experimental error.

2.2.4.5 PCR of *MAT* locus

To determine the mating type at the *MAT* locus we performed colony PCR with forward primers specific for either *MATa* or *MATα* and a common reverse primer (details in Table A.2).

2.2.4.6 Phenotype assays

We assayed ploidy and mating type of the evolved lines by spotting 10 μ L of saturated culture onto agar plates with the two mating type-specific and the diploid-specific media. After two days we noted whether each spot had grown or not. We performed additional mating type assays at the end of the experiment by combining *MAT α* cells with *MAT α* tester strains bearing *his1* and replica plating onto dropout plates (not available for *MAT α* cells because our *MAT α* tester was likely contaminated, see A.4). Because the *his3* auxotrophy in our lines complements the *his1* auxotrophy in the tester strain, only diploids formed by mating can grow on this drop-out medium.

2.2.5 Correlations between traits

To explore correlations between traits (mating rate, cell size, and growth rate), we considered only the lines from the DM treatments. We first determined line means for mating rate in each assay condition, accounting for random effects by finding predicted values from a mixed-effect model. The mating rate measured is shared by the mating types from each evolved line and is used for correlations with both attributes of *MAT α* and *MAT α* cells. We used the mating rate in the mating ratio in which the lines had evolved. For growth rate and cell size we had one measurement per mating type per line. Because one *MAT α* line (168) was aberrant in both mating rate and growth rate (marked with a star in Figure 2.3 and Figure 2.5), we reran our analysis excluding this line, and the results were unchanged. We calculated the Spearman rank correlation for a pair of traits for each group of lines separately and then averaged these correlations across parallel tests to increase power. We tested for statistical significance by randomizing (5000 simulations) the association between trait values within each group to obtain a null distribution for the average rank correlation. We report P-values from two-sided tests.

2.3 Results

Raw data and scripts used for analyses are found at <https://figshare.com/s/94201ff89f5a21ca4f04>.

2.3.1 Mating rate

Mating rate results are shown in Figure 2.2. A linear model including interaction terms between assay ratio and evolutionary treatment outperformed a model without interactions ($\chi^2 = 48.8$, $P < 10^{-5}$), so we made pairwise comparisons between the different mating ratios for each evolutionary treatment group (and ancestor) separately (using the Tukey method for P-value adjustment). These tests were performed to assess whether the evolved lines were better at mating in the mating ratio in which they had evolved. We found no significant differences in the mating rate of the ancestor in the three mating ratios ($P > 0.05$ for all t-ratios between assay conditions). The only evolved group that had a significantly higher mating rate in its evolved condition compared to the other assay conditions was the α -biased lines (t-ratio to unbiased 4.2, to a-biased 9.0, both with $P < 10^{-3}$).

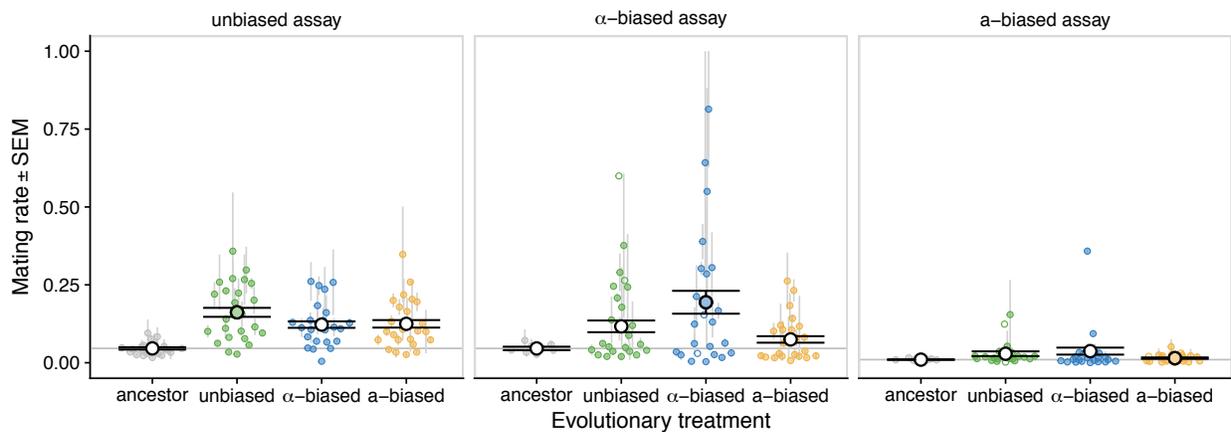


Figure 2.2 Adaptation to breeding system was observed in two out of three evolutionary treatments, based on mating rate of lines from different evolutionary treatments (colours) assayed in different mating-type ratios (panels). In the case where the mating rate of a line was measured only once it is represented by an open circle. When measured more than once, a solid point represents the mean and the grey lines show standard errors. Black circles and error bars show group means and standard errors. The treatment group that evolved in the assayed condition is filled in.

To compare the evolutionary treatment groups with each other and the ancestor, we did pairwise comparisons between the groups separately for each assay condition. When mated in an unbiased ratio, evolved lines of all treatments outperform the ancestor but not one another. The groups have more than, or close to, twice the mating rate compared to the ancestral mating rate of $3.6 \pm 2\%$ (mean \pm SE): $15.5 \pm 2\%$ (t-ratio unbiased – ancestor = 5.38), $11.5 \pm 2\%$ (t-ratio α -biased – ancestor = 3.56), and $11.7 \pm 2\%$ (t-ratio a-biased – ancestor = 3.66) for unbiased, α -biased, and a-biased, respectively ($P < 0.005$ for all comparisons). The only assay condition where one evolutionary treatment showed a significantly higher mating rate than the other two evolved groups was the α -biased treatment when mated with an excess of *MAT α* (t-ratio α -biased – unbiased = 3.0, $P = 0.018$; t-ratio α -biased – a-biased = 4.8, $P < 10^{-4}$), although there was a similar trend with unbiased evolutionary treatment lines performing best in the unbiased assay. No evolved treatment had a significantly higher mating rate than the ancestor or any other treatment group when mated with an excess of *MAT α* cells.

The results of the cross-couple mating rate assays can be seen in Figure 2.3. Similar to the results above, the mating rate in the a-biased ratio was significantly lower than in the α -biased assay ratio ($t = -7.6$, $P < 10^{-5}$). Conditional pairwise tests on the fixed effects of the model revealed that this effect of mating ratio was entirely driven by matings where *MAT α* had evolved as the common partner, which had a mating rate of $58 \pm 5.2\%$ in the α -biased ratio and $16 \pm 5.1\%$ in the a-biased ratio. Regardless of the assay ratio, matings where the *MAT α* line came from the a-biased treatment had significantly lower mating rate ($t = -5.84$, $P < 10^{-5}$), with means of $5 \pm 5.2\%$ in the α -biased ratio and $2.3 \pm 5.1\%$ in the a-biased ratio. Because one *MAT α* line from the α -biased treatment had a very high mating rate compared to the others (line 168), we also ran the model

excluding this one line. Doing so maintained the significant benefit of the *MAT α* partner coming from the α -biased evolutionary treatment ($t = 7.2$, $P < 10^{-5}$), regardless of assay ratio and evolutionary treatment of the *MAT α* partner. It did, however, uncover an additional positive effect: the *MAT α* from the α -biased treatment had a higher mating rate in its evolved condition if its mating partner also came from the same treatment (t -ratio α -biased – a-biased = 3.5, $P = 0.01$). This did not appear to be a specific compensatory effect within coevolved couples, as a likelihood ratio test excluded any significant effect of whether partners came from the same evolutionary replicate ($\chi^2 = 0.02$, $P = 0.88$).

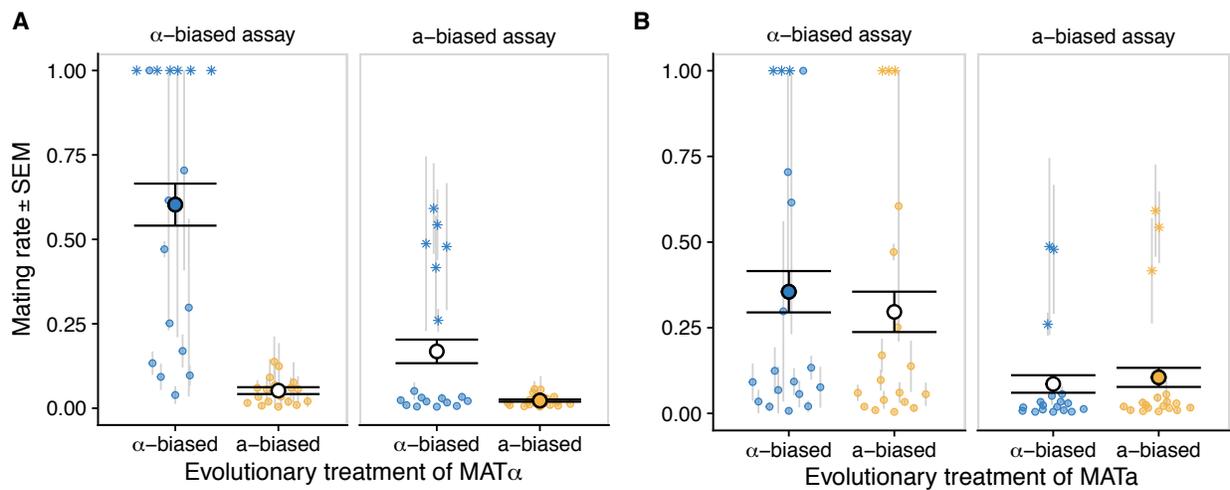


Figure 2.3 Cross-couple mating rates across assay ratio (panels) and evolutionary treatment of the *MAT α* line (A) or *MAT α* line (B) used in the cross. The mating rate of a cross was measured at least twice. The point represents the mean and the grey lines show standard errors. The stars represent crosses to *MAT α* line 168. Black circles and error bars show group means and standard errors. The treatment group that evolved in the assayed condition is filled in.

2.3.2 Cell size and genome content

We assessed both cell size and genome content after the last sexual cycle through flow cytometry. The size of haploid cells decreased for both mating types across all DM treatments when compared to the ancestral haploid lines but not for the IM lines (ANOVA $F = 18.46$, $P < 10^{-5}$, followed by

Tukey test, $P < 0.001$ for all ancestor to treatment group comparisons except for the IM treatment, Figure 2.4A).

We verified the haploid genome content of the lines from the DM treatments. There was no difference in genome content (DNA fluorescence) between the evolved and control haploids (ANOVA $F = 1.066$, $P = 0.37$). While we expected the lines from the IM treatment to be diploid, as they were sampled on day 7 (see Figure 2.1), they too exhibited nearly haploid genome content (Figure 2.4B). Nevertheless, the IM group has a higher genome content than the control haploid and DM lines (ANOVA $F = 19.77$, $P < 10^{-5}$, followed by Tukey test, $P < 10^{-5}$ for all comparisons to the IM group, excluding the diploid control).

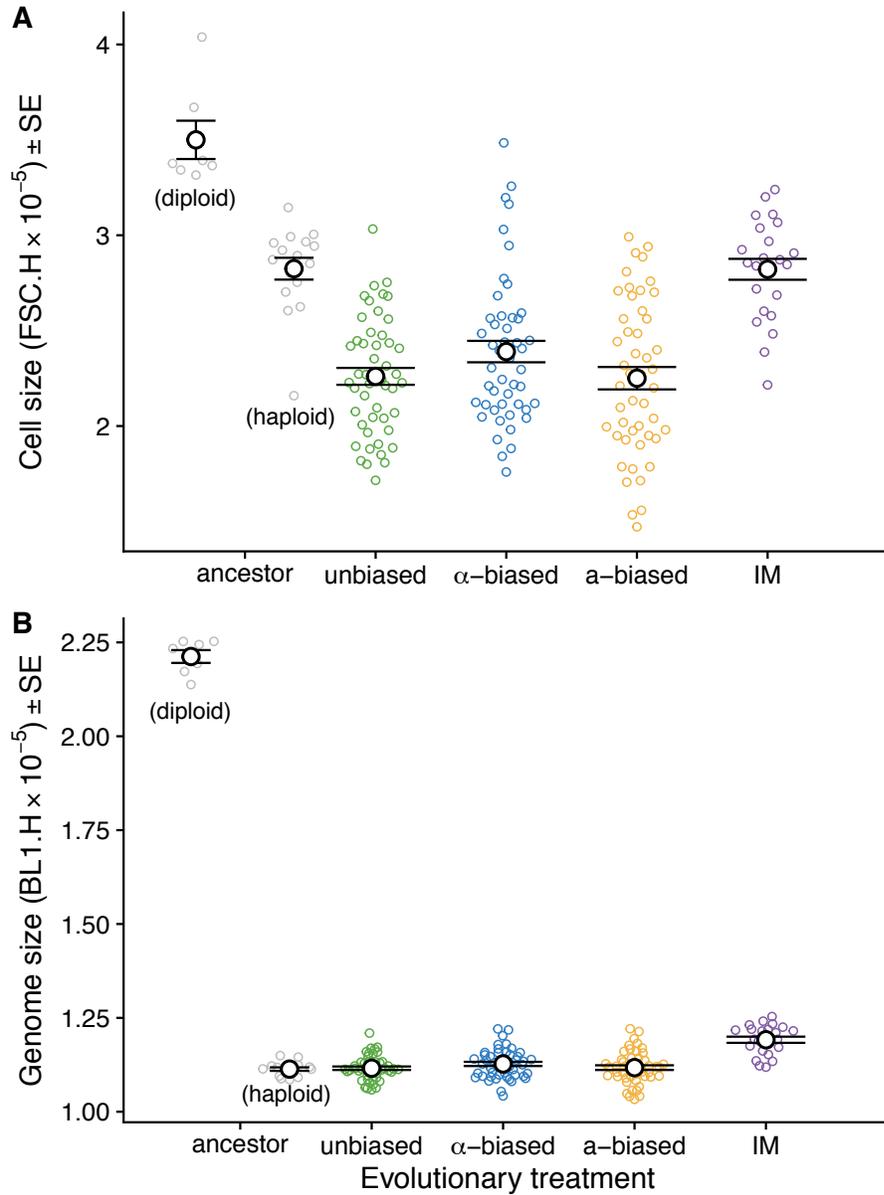


Figure 2.4 Flow cytometry of evolved lines A: Cell size of ancestral diploid and haploid lines as well as evolved haploids as measured by flow cytometry. Cell size decreased in all treatment groups with delayed mating. B: Genome content of ancestors and evolved haploids. Black circles and error bars indicate means and standard errors of treatments.

2.3.3 Phenotype and mating type assay

At the end of the experiment, we spotted the lines after day 7 (haploids) on the haploid selective plates. All of our lines showed the expected pattern of growth, with the *MATa* and *MAT α* lines of

each treatment growing on their mating type-specific medium but not on that of the other. We also mated to tester strains to make sure that the haploids were of the expected mating type. The lines from the IM treatment did not mate with either mating type and exhibited the same growth pattern on the selective plates as our control diploid.

Because the IM lines showed nearly haploid genome content and size while acting diploid in the phenotypic and mating type assay, we assessed their *MAT* genotype through PCR. Each IM line exhibited both *MATa* and *MATα* genes in the active mating type region (Figure A.2). The difference in genome content of the IM lines from the other haploids could potentially be due to aneuploidy (or partial duplication) of chromosome III, where the *MAT* locus is located. This would explain how the IM lines can grow in diploid-selective medium, which requires both mating type markers to be present.

2.3.4 Growth rate

Figure 2.5 shows the relative growth rate of the evolved lines compared to the ancestor. We did not find any significant difference in growth rate between the evolved lines of different treatment groups (at either ploidy level in either permissive or selective medium) and so compared the evolved lines as one group to the ancestor. The evolved *MATa* lines had a significantly higher growth rate compared to ancestral lines in the mating-type selective, but not the permissive, medium ($t = 6.69$, $P < 10^{-5}$ and $t = 0.62$, $P = 0.54$ respectively). The evolved *MATα* lines had a higher growth rate compared to the ancestor in both the mating-type selective and permissive medium ($t = 17.70$, $P < 10^{-5}$ and $t = 3.81$, $P < 0.005$). The evolved diploid lines increased their growth rate in the diploid selective, but not permissive, medium ($t = 3.38$, $P < 10^{-3}$ and $t = 0.36$, $P = 0.72$). The same pattern was observed in the growth rate of haploid (but mating-type

heterozygous) IM lines when compared to the diploid ancestor ($t = 5.40$, $P < 10^{-5}$ and $t = 0.73$, $P = 0.47$).

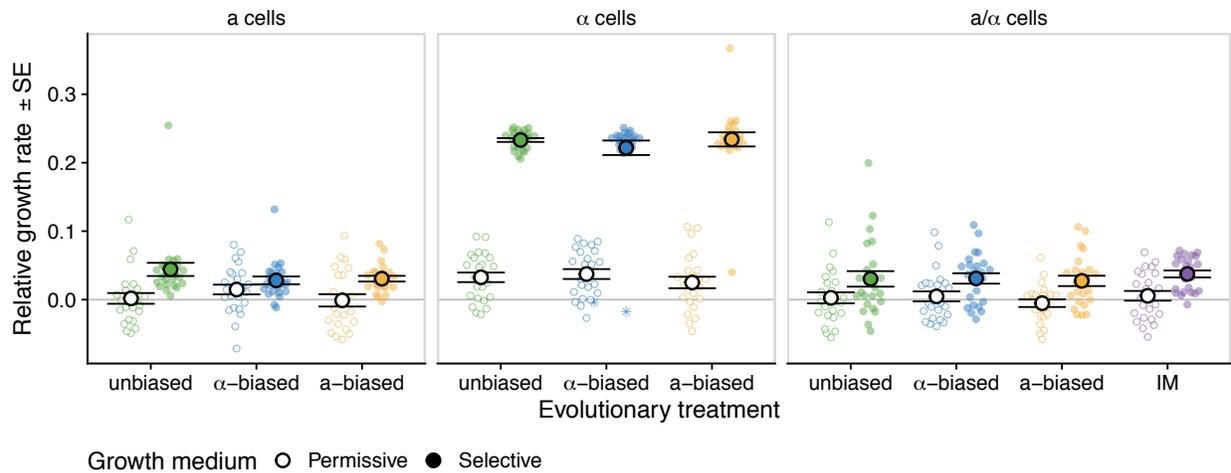


Figure 2.5 Growth rate increased in all selective media. Growth rate of evolved *MATa* cells (left panel), *MATα* cells (middle), and *MATa/α* cells (right) is shown in permissive medium (open circles) or selective medium (e.g., *MATa* haploid selective medium for *MATa* cells in left panel, filled circles). Black circles and error bars show group means and standard errors. Note the low growth of *MATα* line 168 (star)

2.3.5 Sporulation rate

We assayed sporulation rate (measured as the frequency of tetrads) of a sample of populations from each treatment every experimental cycle during the experiment (Figure 2.6A). There was a significant interaction between experimental cycle and the IM treatment ($t = -7.03$, $P < 10^{-5}$). This follows our observation of the IM lines having haploid genome content and a *MATa/α* genotype at the end of the experiment, thus being unable to complete meiosis. Because we had repeated measurements of the same populations, we also evaluated a mixed-effect model that included population ID as a random effect. This model gave the same result but had convergence issues.

After the 10th sexual cycle, we sporulated all diploid lines from the DM treatment, as well as the diploid ancestor (the frozen cultures from day 3 of cycle 1 and of cycle 10). The sporulation rate of the unbiased treatment was significantly lower than the ancestor in our hemocytometer

counts after two days in sporulation medium ($t = -2.06$, $P = 0.04$, Figure 2.6B), with similar trends for all three DM groups.

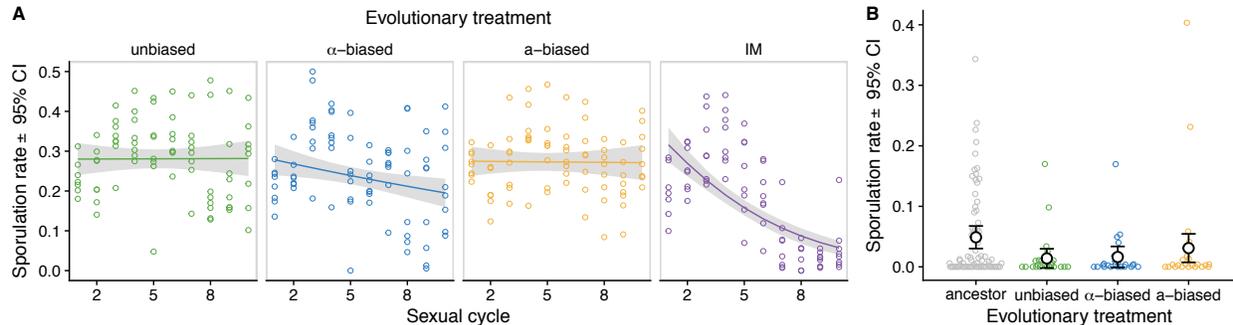


Figure 2.6 Sporulation rate of lines during (A) and after (B) the experiment. Each colored open circle is one hemocytometer count. Curves and grey ribbons in (A) show predicted means (generalized linear model) and 95% confidence intervals. Black circles and error bars in (B) show predicted means by our generalized model and 95 % confidence intervals.

2.3.6 Correlations between traits

We found a significant negative correlation between cell size and growth rate in the selective medium in the DM treatments ($r = -0.195$, $P = 0.023$). In addition, this correlation was significantly more negative in the biased mating ratio treatments compared to the unbiased mating treatment (biased minus unbiased correlation = -0.47 , $P = 0.008$). Interestingly, cell size was not significantly correlated between *MATa* and *MATα* cells across the evolved lines, and the same was true for growth rate. This suggests that the genetic basis for the negative trade-off between cell size and growth rates was, at least in part, mating-type specific.

The average correlation between growth and mating rate (assayed at the mating ratio to which the line evolved) is also negative, with suggestive statistical evidence ($r = -0.17$, $P = 0.05$). We hypothesized that this result may be driven by stronger trade-offs in the common type, which is expected to be under stronger sexual selection. Indeed, the correlation was significantly more negative for the mating type that was common ($r = -0.45$, $P = 0.0016$) than for the rare mating

type ($r = 0.09$, $P = 0.56$) in the biased mating treatments (difference between common and rare correlation coefficient: $P = 0.0112$). There was no significant correlation between growth rate and mating rate in either mating type of the unbiased treatment, and no significant difference in the correlation coefficient of the two mating types in this treatment. We found no significant correlation between mating rate and cell size of either the common or rare mating type.

Finally, we determined whether the growth rate of diploid cells was related to the average cell size of the constituent haploid lines. We considered diploid growth in the permissive environment, which is where initial growth of diploids took place in our experimental protocol. Averaging across all three DM treatments, there is evidence of a positive correlation between haploid size and diploid growth ($r = 0.23$, $P = 0.046$), implying that diploid growth would have declined as the haploid cell size decreased in our experiment (Figure 2.4A). This pattern appears to be driven by the biased treatments ($r = 0.31$, $P = 0.027$), whereas there is no significant correlation in the unbiased treatment ($r = 0.076$, $P = 0.70$).

2.4 Discussion

In this study, we grew yeast populations with frequent rounds of sex under different mating regimes. The mating treatments varied from delayed mating (DM) with either equal proportions of the two mating types (unbiased), to a skewed ratio in favor of one mating type (α -biased or α -biased), to a treatment allowing mating soon after germination (IM). The experimental life cycle allocated similar time for vegetative haploid and for vegetative diploid growth in the DM lines, while being free to evolve in the IM lines (Figure 2.1). The total number of asexual generations was less than ~64 per every sexual cycle: this results in strong selection for sexual characteristics relative to wild yeast (Ruderfer et al. 2006; Magwene et al. 2011).

2.4.1 Mating-type specific adaptation

We found an increase in mating rate in all treatments compared to the ancestor when assayed with an equal ratio of the two mating types, confirming that the response to selection for mating traits was strong in our experiment, with no significant differences between evolutionary treatment groups. In the α -biased mating ratio, on the other hand, the lines that evolved under this ratio mated at a higher rate than the other two treatment groups, a sign of adaptation to breeding system (Figure 2.2). Furthermore, we present evidence that it was the common mating type (in this case *MAT α*) that had evolved to be better at mating (Figure 2.3), consistent with this type experiencing stronger mating competition. In contrast, the mating rate in the a-biased ratio was low in all treatment groups, and we did not find adaptation of the lines in the a-biased treatment to this mating ratio. The common mating type (*MAT a*) did not increase its propensity to mate, even as they experienced similar selection as the *MAT α* lines in the α -biased treatment (if not stronger given the low mating rate of the ancestor in this mating ratio).

Why is it that we find no adaptation of the a-biased populations to their mating treatment? We hypothesize that the nature of Bar1p protease produced by *MAT a* cells could be responsible (Sprague and Herskowitz 1981). This extracellular enzyme digests α factor (the mating pheromone produced by *MAT α* cells). Bar1p ensures that the *MAT a* cell does not initiate mating unless it is in close proximity to a *MAT α* cell, *i.e.* the concentration of α factor is high. When α factor reaches and remains in the pheromone receptors of a *MAT a* cell, the cell stops dividing and prepares for mating. In liquid culture, the ratio of *MAT a* and *MAT α* cells determine the relative concentration of Bar1p and α factor, and hence, the likelihood of mating for the *MAT a* cell (Banderas et al. 2016). We suspect that the abundance of Bar1p in the a-biased mating ratio effectively removed α factor in the culture, allowing *MAT a* cells to continue vegetative growth and further skewing the ratio

toward *MATa*. A mutation to decrease production of Bar1p could not invade, as the effect of a single cell on the amount of enzyme in the medium would be negligible.

In contrast to the Bar1p protease produced by *MATa* cells, the corresponding enzyme in *MAT α* cells that degrades a-factor, Afb1p, is not excreted but rather cell-associated (Steden et al. 1989; Marcus et al. 1991 for suggestive early studies and Huberman and Murray 2013 for isolation, characterization, and mapping of the Afb1 protein). Consequently, mutations in this gene could lead to *MAT α* receiving a-factor and mating earlier than their competitors. The difference in localization between Bar1p and Afb1p could underlie our observation that *MAT α* cells, but not *MATa* cells, evolved to be better competitors at mating when common.

Even though our data cannot tell us the mechanism by which the increased mating rate was reached, we can speculate. Besides Bar1p and Afb1p mutations (discussed above), increased expression of mating-pheromone (to overcome the degradation of Bar1p and Afb1p) and mating-pheromone receptors (to increase sensitivity to the presence of the other mating type) may be favored. It is possible that selection for haploid growth acted in concert with the latter: by design, *LEU2* is under the same promoter as the *STE3* receptor used by *MAT α* , so a change in their transcription factors could have led to both higher growth rate in leucine-limited medium (Figure 2.5) and increased mating rate (Figure 2.2). This cannot be the only explanation for the increase in mating rate, however, because we do not see similar increases in mating rate in the other DM groups (that had similar selection for *MAT α* growth). Furthermore, it fails to explain the negative correlation we found between mating rate and growth rate in the common mating type in the biased mating ratios.

2.4.2 Response in other traits

Although we see dramatic increases in haploid and diploid growth rate, as well as the rate of mating, we observe little or no difference in sporulation rate between the DM evolved and ancestral lines. Because *S. cerevisiae* is known to be good at mating and relatively poor at sporulating (especially lab lines), we expected strong selection for increased sporulation ability, as a “weak link” in the sexual cycle. Studies on both natural isolates (Gerke et al. 2006) and laboratory strains (Codón et al. 1995; Deutschbauer et al. 2002; Enyenihi and Saunders 2003; Deutschbauer and Davis 2005; Ben-Ari et al. 2006; Hall and Joseph 2010) suggest that many loci are involved in determining sporulation efficiency and that mutations arise that both increase and decrease sporulation rate. In addition, increased sporulation rates in response to frequent rounds of sex have been shown to evolve in other experimental studies with yeast (Zeyl et al. 2005b; Thomasson et al. 2020). Interestingly, the one study that evolved lines under mating competition saw decreases in sporulation rate (Reding et al. 2013), suggesting a potential trade-off.

Our correlation analysis suggested two possible trade-offs. First, we observed a trade-off between growth rate (in the haploid-selective medium) and cell size in haploids, which more strongly affects the biased mating ratio treatments. Second, our analysis revealed that mating rate significantly and negatively correlates with growth rate, but only in the common mating type of the biased treatments. These results suggest that two alternative evolutionary strategies were available for the mating type under the strongest sexual selection: either gain a numerical advantage over competitors through increased growth (i.e., those lines with higher growth rates but lower per-capita mating rates and smaller cells) or increase attractiveness to potential mating partners (i.e., those lines with lower growth rates but high per-capita mating rate and larger cells). While the trade-off between mating rate and growth rate has been reported previously (Lang et al.

2009), our study suggests increasing either trait may be a viable response to increased sexual selection. Of these strategies, an increase in gamete numbers at the expense of gamete size appears to result in reduced growth of the diploids that are formed, which is a key trade-off in models of the evolution of anisogamy (Parker et al. 1972; Randerson and Hurst 2001; Parker 2011). Further experimental evolution of yeast under strong sexual selection may therefore shed light on the evolution of sexual size dimorphism.

2.4.3 Haploids with heterozygous mating type escape costs of mating and meiosis

In the IM treatment that selected for mating within a day of germination, all of our lines evolved a haploid genome content while having a heterozygous mating type, allowing them to escape three selective pressures: mating, diploid selection, and meiosis. With a heterozygous mating type, costly haploid-specific expression is eliminated (Lang et al. 2009; Haber 2012), which could give a growth advantage compared to the ancestor during germination (where in addition, haploid growth would be arrested by mating pheromones). Haploids disomic for chromosome III (*MATa/α*) are able to successfully sporulate only if recombination between homologs is inhibited (Wagstaff et al. 1982). Mutations allowing these haploids to form a spore wall could explain their resistance to zymolyase, the digestive enzyme used to kill vegetative cells after sporulation. Our assay for sporulation focused on signs of meiosis and counted tetrads, finding low levels in the IM lines. Future work could distinguish spores, regardless of the number of cells they contain, including using a phase-contrast microscope (e.g. Zeyl et al. 2005) or fluorescent markers (e.g. Gerke et al. 2006).

The evolutionary shift away from diploidy in our IM lines is the reverse of what is commonly observed among asexually propagated haploid lines. These often evolve to become diploid while remaining homozygous at their mating-type locus (Gerstein et al. 2006; Gerstein and

Berman 2015). In fact, our goal for using the immediate mating treatment rather than an asexually propagated haploid control was to avoid diploidization of the latter.

2.5 Conclusion

We forced yeast to go through alternating rounds of asexual and sexual reproduction under different mating systems to understand the evolutionary responses to increased sexual competition. Despite strong selection for meiotic capacity, we observed no evolutionary change in sporulation rate. The trait correlations we observed suggest that the common mating type can either evolve to grow more rapidly, providing a numerical advantage during mate competition at the expense of cell size and diploid fitness, or can evolve to become more attractive to the opposite mating type, despite a slower growth. This finding supports the theory that disruptive selection on gamete traits would lead to anisogamy. We found that *MAT α* cells responded more strongly to sexual selection, evolving to become more efficient at mating, and we hypothesize that this may be caused by biological differences in the way *MAT α* and *MAT a* cells respond to the presence of pheromones of the opposite mating type. This result backs suggestions that asymmetry in gamete signaling is an necessary feature of single-celled eukaryotic mating systems (Hadjivasiliou and Pomiankowski 2016), and offers a different perspective on the evolution of sexual dimorphism. Thus, while isogamous at a morphological level, at molecular and functional levels, *S. cerevisiae* is anisogamous, influencing how it evolves in response to sexual selection.

Chapter 3: From function to fitness

3.1 Introduction

Assume that life is well adapted. Mutations that are deleterious to the function of a protein tend to be eliminated, while neutral variants may rise in frequency. When we compare similar proteins from different species, we find that variation is unequally distributed across the amino acid sequence: some sites are highly variable while other sites are highly constrained. Variation in protein sequences tells us what amino acids can be replaced by each other while maintaining protein function. The assumption is that absence of an amino acid in a particular site in an alignment of homologous proteins signifies a deleterious effect of that amino acid on protein functionality. Because we assume the existing variation has already been optimized by selection, new variants can only either be neutral (if they have high similarity to homologous proteins) or deleterious (if they differ in a substantial way from its conspecifics). By aligning a protein variant of interest to similar proteins scientists can assess the likely functionality of the protein. The philosophical underpinning of this approach is that the observed frequency of a protein variant in homologous proteins is predictive of its effect on fitness.

Several programs have been developed using the approach of sequence alignment described above, and their utility has been validated based on known disease-causing alleles in humans (Ng and Henikoff 2001, 2006). Indeed, protein variants known to cause disease in humans are often poorly represented in homologous proteins of other species. PROVEAN (Protein Variation Effect Analyzer, Choi et al. 2012) has come to be a popular tool in genomic inference and has been used to assess functional differences between proteins in at least 1516 research studies as of October 2020 (Web of Science citation report). Though PROVEAN was developed with and evaluated on human medical data it has increasingly been used to assign functional effects

to identified polymorphisms in a variety of non-human species (Table 3.1). Notably, eco-evolutionary studies have used PROVEAN to find putative causal variants for divergence between populations (Perrier et al. 2017) and species (Wang et al. 2020) that differ not in one, but many, proteins. How well does this tool, which was developed with data from humans, work in predicting fitness effects in other organisms?

Experimental systems offer an exciting opportunity to evaluate the explanatory power of PROVEAN in non-human systems. In particular, mutation-accumulation (MA) experiments attempt to separate mutation from selection by repeatedly bottlenecking populations to small size so that drift rather than selection dominates the probability that a mutation establishes within a lineage (Halligan and Keightley 2009). With a nearly random sample of mutations in a large number of MA lines, we can use PROVEAN to predict the functional effects of the mutations that have accumulated and compare these predictions to direct estimates of relative fitness based on laboratory assays. We can also compare how well the default human-set threshold for sequence and alignment works to find homologous sequences in non-human species.

In this study, we test the explanatory power of PROVEAN scores using growth rate data from three sets of mutation accumulation (MA) lines: *Saccharomyces cerevisiae* dataset 1 (Sc1, Sharp et al. 2018), *S. cerevisiae* dataset 2 (Sc2, Liu and Zhang 2019) and *Chlamydomonas reinhardtii* (Cr, Ness et al. 2015; Kraemer et al. 2017a). These datasets span different kingdoms, ploidy levels (Sc1), environments (Sc2) and genetic backgrounds (Cr). We aimed to determine whether differences in growth rate among MA lines were well predicted by the alignment-based score assigned by PROVEAN. As we show, PROVEAN adds limited predictive power in both the budding yeast datasets and in the *C. reinhardtii* dataset. We argue that there is a need to critically

evaluate the applicability of similar mutational effect-prediction algorithms in the field of ecology and evolution.

Table 3.1 The use of PROVEAN is widespread in ecology and evolutionary studies

Reference	Species	Cut-off score	Number of proteins	Trait/pattern
Takahashi et al., 2017	Experimental mouse	not reported	1	Cataract
Vázquez-Miranda et al., 2017	Le Conte's thrasher <i>Toxostoma lecontei</i>	default	17	Plumage color
Mercatanti, Lodovichi, Cervelli, & Galli, 2017	Yeast <i>Saccharomyces cerevisiae</i>	not reported	NA	Human cancer-associated variants
Gammerdinger, Conte, Sandkam, Penman, & Kocher, 2019	Cichlids Pseudocrenilabrinae	default	>600	Sex-determination
Bajda et al., 2017	Red spider mite <i>Tetranychus urticae</i>	default	1	Resistance to insecticide
Henson et al., 2017	Red wolf and Maned wolf	default	1	Intestinal bowel disease
Cheng et al., 2017	Soybean	not reported	14	Resistance to fungi <i>Fusarium graminearum</i>
Khalid, Khalid, Gul, Amir, & Ahmad, 2018	Common wheat <i>Triticum aestivum</i>	default	1	Drought tolerance
Slugina, Shchennikova, Pishnaya, & Kochieva, 2018	Tomato <i>Solanum lycopersicum</i>	not reported	15 (orthologs)	Self-compatibility
Cabrera, Castellano, Recio, & Alvarez, 2019	Crested wheat grass <i>Agropyron cristatum</i>	not reported	3	Grain hardness
Alavijeh et al., 2020	Citrus red mite <i>Panonychus citri</i>	default	1	Resistance to insecticide
Wang et al., 2020	Fruit bats	-1.3	878 (orthologs)	Convergent evolution of frugivory
Klimushina, Kroupin, Bazhenov, Karlov, & Divashuk, 2020	Common wheat <i>Triticum aestivum</i>	default	7 (orthologs)	Amylose synthesis
Sun et al., 2019	Small brown planthoppers <i>Laodelphax striatellus</i>	default	3	Climate adaptation (cold tolerance)
Alvarez, Castellano, Recio, & Cabrera, 2019	Wild barley <i>Hordeum chilense</i>	default	1	Amylose synthesis
Yoshida et al., 2019	Stickleback <i>Gasterosteus aculeatus</i>	abs(2.5)	1	Hybrid male sterility
Lecová, Tůmová, & Nohýnková, 2019	<i>Giardia intestinalis</i>	default	2	Population structure
Diaz Caballero et al., 2018	Proteobacteria <i>Burkholderia multivorans</i>	not reported	62	Antibiotic resistance

Reference	Species	Cut-off score	Number of proteins	Trait/pattern
Al Khatib, Al Thani, & Yassine, 2018	H1N1 influenza	default	1	Protein evolution
Yoshitomi et al., 2018	Coxsackievirus (CV)-A6	not reported	1	Clinical symptoms
Sharma et al., 2018	HIV-1	default	1	Protein evolution
Kadian et al., 2018	<i>Plasmodium vivax</i>	default	1	Protein evolution
Martino et al., 2018	<i>Lactobacillus plantarum</i>	default	3	Assess function of candidate variants for adaptation
Zepeda Mendoza et al., 2018	Vampire bat <i>Desmodus rotundus</i>	default	At least 4	Assess function of candidate variants for adaptation
Perrier et al., 2017	Lake trout <i>Salvelinus namaycush</i>	default	207	Estimate mutational load within species
Danaher et al., 2017	HSV-1	not reported	64	Find variants associated with plaque size
Yakubu, Salako, De Donato, & Imumorin, 2017	Nigerian goats	default	1	Detect potential sites under positive selection
Kusakabe et al., 2017	Stickleback <i>Gasterosteus aculeatus</i>	abs(2.5)	530	Filter QTL candidates for adaptative differences between ecotypes
Yoshida et al., 2016	Stickleback <i>Gasterosteus aculeatus</i> <i>G. nipponicus</i>	abs(1.3) abs(2.5) abs(4.1)	38	Find functional differences between proteins in two different species
Morales, Pavlova, Joseph, & Sunnucks, 2015	Yellow robin <i>Eopsaltria australis</i>	-4.1	12	Evaluate candidate variants for mitochondrial lineage differences
Hodgins et al., 2015	Asteraceae	default	100s	Estimate mutational load between species
Heuermann et al., 2019	Zea maize	not reported	2	Find causal variants for mutant phenotypes
Fisher, Kryazhimskiy, & Lang, 2019	<i>Saccharomyces cerevisiae</i>	not reported	9	Describe adaptive mutations from experimental evolution
Buhl et al., 2019	<i>Pseudomonas aeruginosa</i>	default	21	Find candidate variant for resistance
del Olmo, Poza-Viejo, Piñeiro, Jarillo, & Crevillén, 2019	<i>Brassica rapa</i>	default	1	Find candidate variants for flowering time
Bhattacharya, Dhar, Banerjee, & Ray, 2019	Common wheat <i>Triticum aestivum</i>	default	1	Assess functional impacts of differences between two consensus domains of protein
Longo et al., 2019	<i>Klebsiella pneumoniae</i>	default	6	Find candidate variant for colistin resistance
Parker, Berny Mier y Teran, Palkovic, Jernstedt, & Gepts, 2020	Common bean <i>Phaseolus vulgaris</i>	default	1	Functional importance of candidate variant for pod indehiscence

Reference	Species	Cut-off score	Number of proteins	Trait/pattern
He et al., 2019	Vampire bats <i>Phyllostomidae</i>	default	1	Find candidate variants for immunesystem adaptation
Ofori-Anyinam et al., 2020	<i>Mycobacterium africanum</i> <i>Mycobacterium tuberculosis</i>	default	18	Assess candidate variants between lineages for metabolic differences
Alfano et al., 2020	Tick-borne encephalitis (TBE)	default	1	Assess variants compared to reference strain
Tancos et al., 2020	<i>Rathayibacter toxicus</i>	default	20	Assess functionality of antibiotic genes
Seung, Echevarría-Poza, Steuernagel, & Smith, 2020	<i>Arabidopsis thaliana</i>	not reported	1	Filter sequence variants for functional effects
Bazhenov, Chernook, Kroupin, Karlov, & Divashuk, 2020	Wild grass <i>Dasypyrum villosum</i>	not reported	1	Assess allelic diveristy
Rowland et al., 2020	Tomato <i>Solanum lycopersicum</i>	default	1	Find variants for trait variation
Piombo, Bosio, Acquadro, Abbruscato, & Spadaro, 2020	Hemibiotrophic fungus <i>Fusarium fujikuroi</i>	not reported	36	Find variants for trait variation
Stage et al., 2020	<i>Lactobacillus rhamnosus</i>	not reported	6	Assess phenotypic impact of SNPs
Castellá, Bragulat, Cigliano, & Cabañes, 2020	<i>Aspergillus carbonarius</i>	not reported	1	Assess variation within species
Veilleux, Louis, & Bolnick, 2013	Nocturnal lemurs	not reported	1	Assess variation among species; estimate mutational load
Schweizer et al., 2016	Grey wolves	not reported	1	Functional impact on protein
Renaut & Rieseberg, 2015	Compositae crops	default	>1000?	Estimate mutational load in domesticated species
Conte et al., 2017	Spruce	default (but tests others)	6928	Estimate mutation load
Li et al., 2014	Antarctic penguin	default	4922	Find candidate genes for adaptation
Harrison, Guymer, Spiers, Paterson, & Brockhurst, 2015	<i>Pseudomonas fluorescens</i>	not reported	2	Find adaptative variants from experimental evolution
Rasal et al., 2016	Zebrafish <i>Danio rerio</i>	default	1	Single gene analysis
Zhang, Zhou, Bawa, Suren, & Holliday, 2016	Black cottonwood <i>Populus trichocarpa</i>	not reported	~20000?	Assess mutational load
Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016	Review	na	na	Assess inbreeding depression

Reference	Species	Cut-off score	Number of proteins	Trait/pattern
Yoshida, Makino, & Kitano, 2017	Japan Sea Stickleback <i>Gasterosteus nipponicus</i>	default	1280	Assess mutational load
Lind, Arvidsson, Berg, & Andersson, 2017	<i>Salmonella typhimurium</i>	not reported	3	Classifying mutational effects of proteins in bacteria and relating to competitive fitness
Liu, Zhou, Morrell, & Gaut, 2017	Rice <i>Oryza sativa</i>	not reported	thousands	Assess mutational load
Ochoa, Onorato, Fitak, Roelke-Parker, & Culver, 2017	Florida panthers and Texas puma	default	2	Screening for mutations deleterious for species
Gorter et al., 2017	Yeast <i>Saccharomyces cerevisiae</i>	default	~200	Assess adaptations under experimental evolution
Navarro-Sigüenza, Vázquez-Miranda, Hernández-Alonso, García-Trejo, & Sánchez-González, 2017	Woodpecker <i>Melanerpes</i>	default	6	Assess variation among species
De Lomana et al., 2017	yeast	not reported	2	Assess adaptations under experimental evolution
Makino et al., 2018	Multiple domesticated species	default	Tens of thousands	Assess mutational load
Priyam, Tripathy, Rai, & Ghorai, 2018	Multiple reptilian species	default	4	Assess variation among species
Ferchaud, Laporte, Perrier, & Bernatchez, 2018	Lake trout <i>Salvelinus namaycush</i>	default	~4000	Assess mutational load
Al-Shuhaib et al., 2018	Ostriches	default	6	Find candidate loci for adaptation
Hall, Harrison, & Brockhurst, 2018	<i>Pseudomonas</i> spp.	not reported	1	Assess variation among species; candidate variants for trait variation
Hamabata et al., 2019	Endangered island endemic plants	abs(2.5)	Tens of thousands	Assess mutational load
Yoshida et al., 2020	Threespine stickleback <i>Gasterosteus aculeatus</i>	default	3512	Assess mutational load

3.2 Background

Predicting the impact of amino acid substitutions on protein function has been a long-standing goal of molecular evolutionary research. An early approach compared proteins among related species to determine which amino acids were most substitutable over evolutionary time. Highly similar proteins were gathered, and the frequency of different amino acid substitutions in these sequences were used to determine a substitution matrix. Henikoff and Henikoff (1992) gathered conserved regions of proteins (“blocks”) from the PROSITE database for over 500 protein families to generate the BLOSUM matrix (short for Blocks substitution matrix). The score for a certain amino acid substitution in the BLOSUM is a log-odds ratio, $\text{Log}(H_1/H_0)$, where H_0 is the probability of seeing the two amino acids align by chance (frequency of amino acid 1 times the frequency of amino acid 2), and H_1 is the frequency of this amino acid pair in the conserved blocks (for a primer to the BLOSUM matrices, see Eddy 2004). The BLOSUM62 matrix that is used for protein alignment in PROVEAN has blocks aligned from proteins that are less than 62% identical. A 62% cut-off ensures that the proteins that are being compared are fairly divergent; within these proteins, only the conserved regions are used in the BLOSUM matrix, ensuring that their similarities reflect selection.

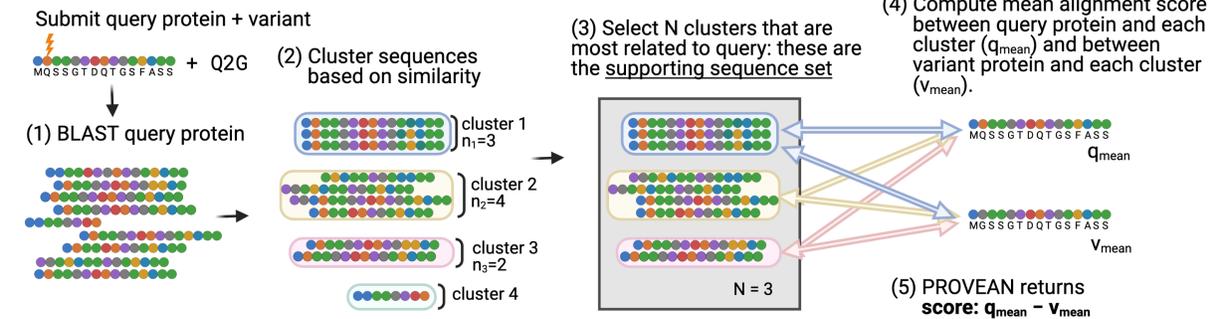
Cargill et al. (1999) used the BLOSUM62 matrix to rank single nucleotide polymorphisms in human proteins as conserved or not. This spurred Ng and Henikoff (2001) to develop SIFT (Separate Intolerant From Tolerant): they argue that the functional impact of an amino acid substitution cannot be solely predicted by its score in the BLOSUM matrix. Rather, one needs to consider the position in the sequence and existing polymorphisms in highly homologous sequences. While BLOSUM62 is based on highly diverged proteins (and the conserved regions thereof), SIFT relies on homologous proteins to predict whether a mutation exists in a functionally

important part of the protein. It calculates a matrix giving the probability of a substitution for each position in the protein. For any given mutant protein, SIFT can predict whether it will be functionally deleterious given its probability in the site-specific matrix.

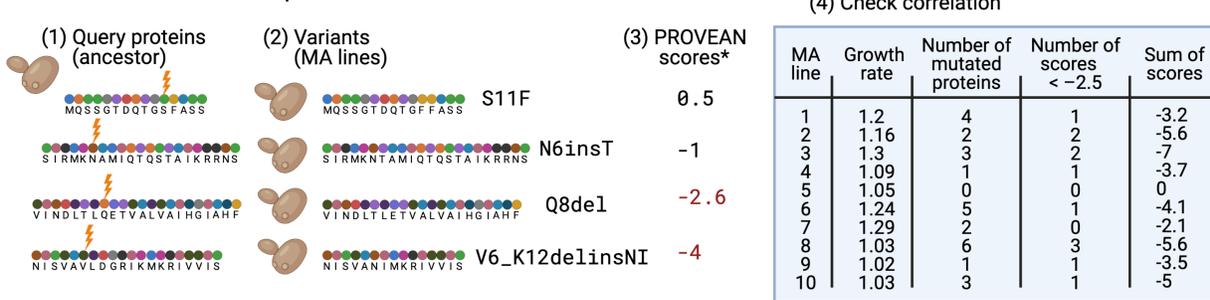
The Protein Variant Effect Analyzer (PROVEAN) was introduced by Choi et al. (2012) as a tool to predict the effects of not only amino acid substitutions but also in-frame insertion and deletions. Similar to SIFT, PROVEAN gathers clusters of highly homologous sequences in the NCBI non redundant protein sequences (nr) database. Rather than producing probabilities of substitution across the protein of interest, PROVEAN computes an alignment score for both the query sequence (*i.e.* wild type) and the mutant to these sequence clusters. The difference in the mean alignment score for the query and mutant protein is called the PROVEAN score.

In the first step of PROVEAN, a BLAST search is done using the submitted query sequence. An Expect (E) value cut-off of 0.1 is used to find homologous, while still distantly related, sequences. This typically results in thousands of matches across a wide range of taxa. To avoid redundancy the found sequences are clustered based on a cut-off of 75% sequence similarity within a cluster. Then, the top 30 clusters most similar to the query sequence are used to calculate alignment scores to the query and mutant sequence and finally to compute difference in mean alignment scores of the query and mutant sequences, the PROVEAN score. The supporting sequence set can be saved and analyzed independently. The PROVEAN process is summarized in Figure 3.1A.

A How does PROVEAN work?



B Can PROVEAN score predict fitness?



C How does PROVEAN score variants between reference genome and lab strain (MA ancestor)? Does it depend on which protein is submitted as the query?

Reference protein	Lab strain protein	Variant (if reference is used as query)	Score	Variant (if lab strain is used as query)	Score	Difference in absolute score
MQSSVYFDQTGSFASS	MQSSGT DQTGS FASS	V5_D8delinsGT	-3.7	G5_T6delinsVYF	3.7	0
SIRMKNAMIQSTAINS	SIRMKNAMIQSTAIKRRNS	I16insKRR	2.2	K17_R19del	-2.2	0
VINDLTLQETVALVAIHGIAHF	VINDLTLQETVALVAIHGIAHF	F7L	2.8	L7F	-2.8	0
NISVAVLDRGRQVLTKMKRIVVIS	NISVAVLDRGRQVLTKMKRIVVIS	Q11_L13del	-0.3	R10insQVL	0.7	0.4

Figure 3.1 Conceptual figure of project. (A) Outline of PROVEAN process, inspired by Figure 10 in (Choi et al. 2012). In the final step, each sequence of the supporting set is aligned to the query and variant protein. The mean alignment score of each cluster to query or variant is taken. The average of these average alignment scores makes out the q_{mean} and v_{mean} . (B) All protein variants between the MA ancestor and MA lines were submitted to PROVEAN. (C) We also assess how variants between the reference and MA ancestor are scored. Created with BioRender.com.

As described above, PROVEAN reports the difference in average alignment scores (called PROVEAN score) of the query and mutant protein to the gathered sequence clusters, together with a predicted functional category (either deleterious or neutral). Though it is possible for a mutant protein to receive a higher mean alignment score than the wild type, there is no category for beneficial effects. Whether PROVEAN will call the variant deleterious or not depends on a preset threshold. The default value of this score is -2.5; a PROVEAN score below this is called deleterious. The default cut-off value was determined by maximizing sensitivity (detection) and specificity (accuracy) when assigning functional effects to human protein variants that are common polymorphisms versus disease-causing (Choi et al. 2012; Choi and Chan 2015). The researchers suggest changing the cut-off depending on the user's needs. In the studies summarized in Table 3.1 that applied PROVEAN to non-human organisms, however, nearly all used the default cut-off value. As far as we can tell, there has been no attempt to find other species-specific thresholds. Many previous studies used PROVEAN to draw conclusions about the relative mutational load in a variety of situations: in different habitats of nocturnal lemurs (Veilleux et al. 2013), between native and invasive species of Asteraceae (Hodgins et al. 2015), between domesticated and wild Compositae crops (Renaut and Rieseberg 2015), populations of Lake Trout (Perrier et al. 2017), and between new sex chromosomes in cichlids (Gammerdinger et al. 2019). It is worth noting that PROVEAN operates on the null hypothesis that all large changes to proteins are deleterious (the more dissimilar to the query sequence the worse). Adaptation to a new environment requiring large changes to a protein will thus be misclassified as deleterious. Some studies (*e.g.* Yoshida et al. 2016; Kusakabe et al. 2017) have used the absolute score of PROVEAN to find function altering proteins among populations. This assumes that a protein variant will be scored symmetrically around 0.

Here we analyze three MA datasets that combine genomic information with growth rate data, allowing us to evaluate whether mutations scored as deleterious by PROVEAN correlate with measured fitness (Figure 3.1B).

3.3 Methods

3.3.1 Data processing

3.3.1.1 *S. cerevisiae* dataset 1

This experiment is described in detail in Sharp et al. (2018). In short, 220 replicate yeast lines were propagated under relaxed selection for 100 days of continuous growth (~1500 generations per line). The 220 lines were equally distributed among four groups; haploid and diploid lines had either their copy of the *RDH54* gene (responsible for homologous recombination and repair during mitosis) deleted or intact. At the end of the experiment, the growth rate of 218 of these lines, and 182 replicates of their ancestors (equally distributed among the four groups), were measured in two Bioscreen C machines. The growth rate assay was done 11 times, over consecutive days. The growth rates from one machine on one day was omitted from the data analysis because of a shaking malfunction, such that the number of observations varies between 10 and 11 for the lines. There was a reduction in the mean growth rate of diploid, but not haploid, MA lines compared to controls. Nevertheless, there was significant genetic variation in all groups of MA lines, but in no group of the control lines (evaluated by likelihood ratio test of mixed effect model with or without line identity as an explanatory factor, Sharp et al. 2018). The genomic analysis revealed high rates of aneuploidy in the diploid lines, which could explain a large part, but not all, of the reduction in fitness of the diploid MA lines.

We used the mutations reported in Sharp et al. (2018; Dataset_S2.xlsx). There were 1474 genic mutations in the dataset, occurring in 1219 unique genes across 218 MA lines. We extracted

the nucleotide and protein sequence of the genes affected using YeastMine. From the same database, we downloaded the location of introns in these genes. The reference nucleotide sequence was then mutated *in silico* to represent the mutant sequence, which was then transcribed and translated, using the seqinr package in R. Additionally, we analyzed VCF files to obtain a table of mutations in the ancestral line as compared to the yeast reference genome (version R64-2-1). In cases where the ancestor and reference strain differed for a mutated gene (126 genes) we separately computed the ancestral protein and used it for comparison to the MA lines. We wrote an algorithm to produce protein variants in the format PROVEAN requires. From 1474 genic mutations, 1126 protein variants were computed (in 961 unique proteins). Three samples (lines 113, 202, and 206) had no nonsynonymous mutations and were dropped. When an MA line had more than one nonsynonymous mutation in a particular gene both mutations were considered when altering the protein and the number of mutant proteins is reported once.

The growth rates reported in Sharp et al. (2018; Dataset_S2.xlsx) include all replicate measurements for each line. We evaluated the percentage of variation in growth rate explained by line identity in the ancestor and MA lines. In the control group, where each line was an isogenic copy of the ancestor, 2% of variance could be attributable to line identity. In the MA lines, 15% of variance was attributable to differences among lines. The intra class correlation for MA line was 0.169 (computed with the performance package in R). Unless otherwise noted, we used all replicate measurements of each MA line to create one response variable that we call relative fitness: the average growth rate of each MA line compared to the average growth rate of the ancestral line with the same mating type, ploidy (haploid or diploid), and *RDH54* gene status (presence or knockout of gene involved in homologous repair), accounting for random block effects (for the significance of mating type and *RDH54* status refer to Sharp et al. 2018). There

were two lines for which there was no growth rate data; these lacked mitochondrial function (“petite”) by the end of the MA experiment and were excluded from the analysis (lines 110 and 189). Information on sequence depth, heterozygosity of mutations in diploids, and changes in chromosome copy number were compiled by N.S.

3.3.1.2 *S. cerevisiae* dataset 2

This experiment is described in detail in Liu & Zhang (2019). In short, 168 replicate diploid yeast lines were propagated under relaxed selection for around 60 bottlenecks under continuous growth (~1000 generations per line). The 168 lines were equally distributed among seven treatment groups that were grown in different kinds of media. Two lines were lost during MA, likely due to lethal mutations. At the end of the experiment, the growth rate of frozen samples from the beginning of the experiment and evolved MA lines were measured through cell counts of single colonies (one colony per line) on the same medium as they had evolved. There was a general reduction in mean growth rate of the MA lines compared to controls. The genomic analysis revealed high rates of aneuploidy and segmental duplications in two environments, which could explain a large part, but not all, of the reduction in fitness.

We used the mutations reported in Liu and Zhang (2019; Data_S1.xlsx). Additionally, Dr. Zhang supplied us with a table of mutations in their ancestral line relative to the S288C reference genome. We used the same method as described above. There were 1147 genic mutations, occurring in 968 unique genes, across 165 MA lines. From 1147 genic mutations, 877 protein variants were computed (in 754 unique proteins). Out of 754 altered proteins, 16 already differed between the S288C reference genome and the laboratory ancestor, in which case the latter was used as the query sequence.

Growth rates in five different media were reported in Liu and Zhang (2019). Because each line was measured only once, we cannot produce a repeatability estimate. However, we can see that the variance among measurements of the ancestors (23-24 measurements in each condition) is smaller than the variance among MA lines (F-value = 28.23, $P < 10^{-5}$). We computed the relative fitness of each evolved MA line by subtracting the mean initial growth rate of the starting strains in the focal medium from the final growth rate of each line in its medium.

In addition, we used the information on chromosome aneuploidies and segmental duplications and deletions from Liu and Zhang (2019; TableS2.xlsx) to compute a relative genome size for each strain.

3.3.1.3 *C. reinhardtii* dataset

This experiment is described in detail in Morgan, Ness, Keightley, & Colegrave (2014). In short, 15 replicate lines from each of six different *C. reinhardtii* strains (for a total of 90 MA lines) were propagated under relaxed selection for around 85 bottlenecks under continuous growth (~1000 generations per line). At the end of the experiment the experimenters measured growth rates of replicates of the ancestor and MA lines and found an increase in variance in fitness as well as a reduction in mean fitness in the MA lines. The genomic analysis of these lines was presented in Ness et al. (2015). Competitive growth rates of the MA lines and ancestors were reported in Kraemer et al. (2017), where they found significant correlations between the number of nonsynonymous mutations in coding regions and the growth rates of MA lines.

We received an annotated table of the mutations reported in Ness et al. (2015) as well as VCF files containing the mutations in their six ancestral lines compared to the genome reference. We downloaded an annotated table for all transcripts in the *Chlamydomonas* reference genome from Dicots PLAZA 4.0 (version 5.5, Van Bel et al., 2018) to choose mutations that occur in

coding sequences. Out of the original 6843 mutations, 6209 were genic, and 3900 occurred in coding sequences. In the category of coding sequence mutations, as defined by Dicots PLAZA, four were found in a UTR, three were intronic, and four were found in chloroplast or mitochondrial DNA for which we did not have transcript sequences. Removing these, we were left with 3889 mutations, representing 1439 mutated proteins after combining mutations. We found that the majority of transcripts that were mutated during mutation accumulation already had existing variants in the ancestral strain, relative to the reference (Table 3.2). 1397 out of the originally predicted 1439 protein variants remained once ancestral variation had been considered (Table 3.2).

Table 3.2 Number of protein-coding mutations for MA lines of a particular ancestral *C. reinhardtii* line and the ancestral line's protein-coding mutations.

Ancestor strain ID	Number of mutated proteins in MA lines^a	Number of which had prior mutation(s) in ancestor	Number of MA lines	Number of protein variants^b
1373	381	234	12	390
1952	80	75	14	82
2342	188	171	12	192
2344	239	225	15	244
2931	351	322	14	358
2937	130	109	15	131
Total	1369	1136	82	1397

^a Each protein is counted only once, even if several MA lines have mutations in the protein.

^b Each change to a protein in any mutant sample is counted. This is the number of variants submitted to be scored by PROVEAN.

We found 2 cases in the *C. reinhardtii* dataset where the reported reference nucleotide deviated from that found in the Dicots PLAZA 4.0 sequence, in each case, the differences between the two reference sequences were synonymous. This discrepancy was likely due to the two different reference genomes used (Ness et al. used v5.3; Van Bel et al. used v5.5).

The ancestral nucleotide sequence was mutated *in silico* to represent the mutant sequence. The mutated sequence was then translated using the *seqinr* package in R. To test the accuracy of our sequence-mutating code, we mutated the coding sequence to the reference nucleotide given by the *C. reinhardtii* dataset and verified that this produced the reference transcript. We converted the protein variants into the format PROVEAN requires. In cases with alternative transcripts, we treated these as separate proteins in PROVEAN and then report the minimum score given to any protein variant of a gene. This occurred in 42 unique cases, involving all ancestral backgrounds. While the difference in scores between transcripts in general was small, we found two cases where the score for one affected transcript was below the default -2.5 threshold while the other was above it, and six cases where the scores fell above and below zero. Six out of the total 1397 protein variants failed to receive a score from PROVEAN, likely because the changes to the protein were too large to compute alignment scores between the clusters gathered and the mutant protein and were ignored in the analysis (these occurred in six different samples across five ancestral backgrounds).

Growth rates for each MA line were found in the DRYAD repository for Kraemer et al. (2017b). We followed Kraemer et al. (2017a) and excluded three lines for which they estimated extreme mutation rate values; 72 MA lines remained. Because the researchers did not find an effect of treatment on competitive fitness of their lines (Table 3, Kraemer et al., 2017a) we used the estimated fitness in the benign environment for our analysis. Kraemer et al. (2017b) report competitive fitness (calculated as the growth rate of the focal strain minus the growth rate of the competitor strain) for all replicate measurements for each MA line and ancestor. The percentage of variance in competitive fitness attributed to line identity in the MA lines was 39%. The intra class correlation for line identity was 0.40 (computed with the performance package in R). We

used all replicates to create one response variable that we refer to as relative competitive fitness: the average growth rate of each MA line compared to the average growth rate of its ancestor, accounting for the random block effect of plate.

3.3.2 Running PROVEAN

We ran PROVEAN on the ComputeCanada cluster. As the program failed to run with the recent BLAST software (version 2.9.0), we configured PROVEAN to run with psiblast and blastdbcmd from BLAST version 2.4.0. We used version 4.8.1 of CD-HIT. We ran our variants with the NCBI nr database from 12/11/2019, which holds 142 GB of non-redundant sequences (229,636,095 sequences). We ran a subset of variants using the 2012 database, on which PROVEAN was developed (the first 5 GB), without radical changes to the PROVEAN scores of variants. The supporting sequence sets used to compute the alignment scores for all proteins were saved.

We supplied the protein of the ancestor as the query sequence. It should be noted, that PROVEAN has not been validated for frameshift and/or nonsense mutations. We have several protein variants that cause large changes to the protein. These all received very negative scores (Table 3.3). A small set of large protein changes also received large positive scores (which PROVEAN calls neutral). To avoid undue influence of these very negative scores in downstream analyses, we set their value to the lowest (or highest in the case of large positive values) score given to any single amino acid substitution in the data set (-14 for both *S. cerevisiae* datasets, -12.8 and +5.18 for Cr).

Table 3.3 Summary of protein mutations and scores in (a) Sc1, (b) Sc2, (c) Cr1 dataset. Complex mutations refer to cases where more than one kind of mutation (substitution, insertion, or deletion) occur in the same protein.

(a)	Type	Number	Max score	Min score
	single aa substitution	1015	5.22	-14
	duplication	5	0.574	-3.74
	deletion	71	0.813	-4000
	insertion	2	1.64	-4.78
	complex	37	-0.747	-2650
(b)	single aa substitution	723	7.37	-14.0
	duplication	28	2.08	-9.54
	deletion	63	-1.48	-5978
	insertion	1	-0.169	-0.169
	complex	62	4.07	-2911
(c)	single aa substitution	1263	5.18	-12.7
	duplication	27	1.5	-6.84
	deletion	73	0.581	-4259
	insertion	19	1.84	-8
	complex	142	1343	-6620

3.3.3 Statistical analyses

In our analysis, we use a number of summary statistics. The first is the number of mutant proteins per line, Ψ_{total} . Based on the PROVEAN score, each of the protein variants is classified as neutral or deleterious. The second summary statistic is the number of mutant proteins classified as deleterious using the default cut-off of -2.5 by PROVEAN, Ψ_{del} . Because some studies have used the absolute departure from zero (Yoshida et al. 2016, 2019; Kusakabe et al. 2017), we also investigate the number of protein variants scored below -2.5 or above 2.5, Ψ_{abs} . We also investigated whether there is any quantitative information in the PROVEAN score, above the categorical information (neutral or deleterious). Therefore, we explored models using the sum of PROVEAN scores: \sum_{tot} is the sum of mutant protein scores for each MA line.

Initial analyses failed to find a significant effect of the number of altered proteins, Ψ_{total} , on the relative fitness of the yeast lines. In an attempt to account for the importance of a protein to fitness, we used a previously published list of essential genes in yeast (Winzeler et al. 1999, downloaded from the Yeast Deletion Web Pages) and added the “essentialness” of a protein as an explanatory factor in our models of the yeast datasets. Summary variables by the essentialness of the proteins are labelled $\Psi_{\text{tot_ess}}$, $\Psi_{\text{del_ess}}$, $\Psi_{\text{abs_ess}}$, and $\sum_{\text{tot_ess}}$.

To compare the usefulness of our explanatory variables we make a null model for each dataset that includes no information about protein changes. We compare models using the four (or eight in the case of yeast data) different summary variables (Ψ_{total} , Ψ_{del} , Ψ_{abs} , and \sum_{tot}) to the null model using the R package AICcmodavg. We use this package to produce model selection tables based on second order Akaike’s information criterion (AIC).

3.3.3.1 *S. cerevisiae* dataset 1

We analyzed haploids and diploids first together and then separately, given that mutational effects may differ between them due to masking in diploids. For both sets *RDH54* gene status was also included as a predictor in the models. In the diploids, relative genome size was also included. Because haploid MA lines did not experience an overall decrease in growth rate (but still showed genetic variance in fitness), we explored models with the absolute value of relative fitness (a decrease and increase in fitness of the same magnitude will count the same). 18% of the protein variants occurred in essential genes (197 out of 1130).

3.3.3.2 *S. cerevisiae* dataset 2

We used linear models of relative fitness as a function of each PROVEAN predictor, relative genome size (computed by adding or subtracting potentially lost or gained chromosomes or large segmental deletions or insertions), and environment during MA. We combined our analysis of the

strains from the different backgrounds as we found no significant effect of environment on the effect of MA ($F = 1.38$, $P = 0.22$). 20% of the protein variants occurred in essential genes (178 out of 877).

3.3.3.3 *C. reinhardtii* dataset

We used linear models of relative fitness as a function of each PROVEAN predictor and included ancestral background as a random effect. Because competitive fitness only decreased significantly after MA in two out of six ancestral backgrounds (CC2342 and CC2931), we also explored models with the absolute value of competitive fitness as the predictor variable.

3.4 Results

The result files and supporting sequence sets produced by PROVEAN, as well as scripts used for analyses are available for download at https://zoology.ubc.ca/~sandell/provean_project/. As shown below, we find that PROVEAN does not consistently predict fitness effects in the mutation accumulation studies examined when comparing mutant proteins to the proteins of their most immediate laboratory ancestor.

3.4.1 *S. cerevisiae* dataset 1

We evaluated linear models of relative fitness with ploidy, *RDH54* gene status, and relative genome size as fixed effects. We found no predictive effect of the number of mutant proteins Ψ_{tot} (t value = -0.18, $P = 0.86$), the number of deleterious mutant proteins Ψ_{del} (t value = -1.1, $P = 0.30$), the number of mutant proteins with absolute score greater than 2.5 Ψ_{abs} (t value = -0.84, $P = 0.40$), or the total sum of PROVEAN scores Σ_{tot} (t value = 0.96, $P = 0.304$). Model comparisons show that none of the models evaluated had a lower AICc than the null model (with the fixed effects of ploidy, *RDH54* gene status, and genome size). We received the same qualitative result for all tests

and model comparisons when evaluating the predictive power of changes only to essential proteins (see Figure 3.2).

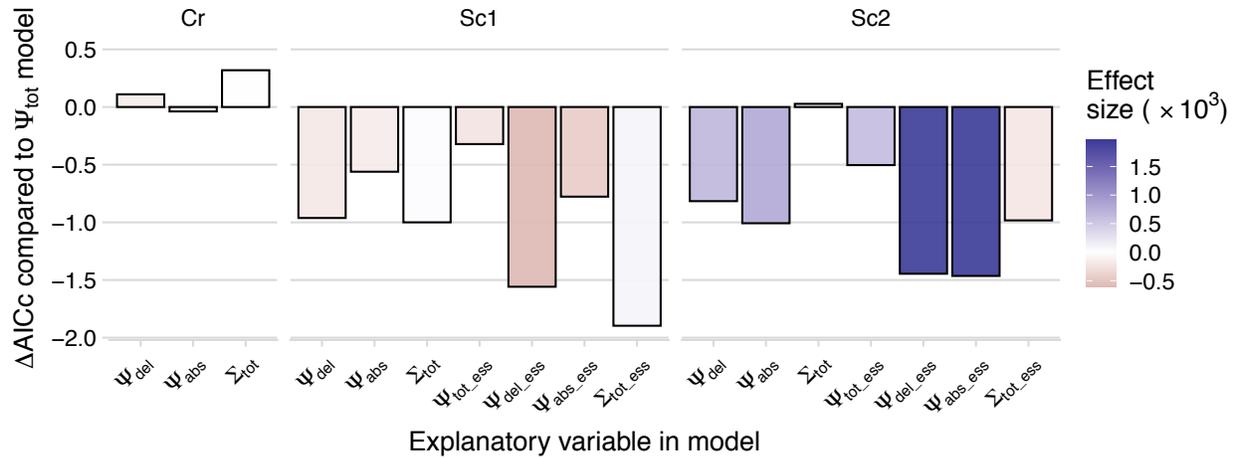


Figure 3.2 PROVEAN prediction adds little explanatory power to growth rate differences among MA lines

We proceeded by separating the dataset into haploid and diploid MA lines. Running the same set of models as above but using only diploid lines (such that the only explanatory variable in the null model is *RDH54* gene status and relative genome size) we got the same qualitative result as in the whole dataset. In the haploid lines, we focus on the absolute value of relative growth rate as the predictor variable, as the lines did not decline significantly in mean fitness. Because there were no large changes in genome size in the haploid lines, the null model for this set is fitted with only *RDH54* status as the explanatory variable. In addition, there was no score above +2.5 in the haploid lines (meaning $\Psi_{del} = \Psi_{abs}$ and $\Psi_{del_ess} = \Psi_{abs_ess}$). We found no predictive effect of the number of mutant proteins, whether considering all or only essential proteins: Ψ_{tot} (t value = 1.4, P = 0.18), Ψ_{del} (t value = 1.4, P = 0.18), Ψ_{tot_ess} (t value = 0.25, P = 0.80), Ψ_{del_ess} (t value = -1.2, P = 0.24). We found that the sum of PROVEAN scores Σ_{tot} had a significant negative effect on the absolute change in relative growth rate (t value = -2.1, P = 0.04), but this result is opposite of what would

be expected if PROVEAN scores accurately predicted change in fitness. When considering all mutant proteins in haploids, the Σ_{tot} model outperforms the null with a ΔAICc of -2.21, but does not when considering only essential proteins (t value = 0.13, P = 0.89, the null model again was the model of choice).

3.4.2 *S. cerevisiae* dataset 2

We evaluated linear models of relative fitness with MA environment and relative genome size as fixed effects. We found no predictive effect of the number of mutant proteins Ψ_{tot} (t value = 0.70, P = 0.49), the number of deleterious mutant proteins Ψ_{del} (t value = 0.81, P = 0.42), the number of mutant proteins with absolute score bigger than 2.5 Ψ_{abs} (t value = 0.87, P = 0.39), or the total sum of PROVEAN scores Σ_{tot} (t value = -0.62, P = 0.53). Notice also that the effect of these variables has opposite sign compared to their effect in Sc1 above. Model comparisons show that none of the models evaluated had a substantially lower AICc than the null model (with the fixed effects of genome size and MA environment) (Figure 3.2). None of the P values is small enough to withstand any correction for multiple tests.

We then turned to models using changes only to essential proteins. We found a suggestive effect of the total number of essential proteins mutated, $\Psi_{\text{tot_ess}}$ (t value = 1.737, P = 0.08) and a significant positive effect on growth rate of the number of essential proteins with deleterious scores $\Psi_{\text{del_ess}}$ (t value = 1.995, P = 0.048). The effect was very similar when considering the absolute value of PROVEAN score, $\Psi_{\text{abs_ess}}$ (t value = 1.985, P = 0.049). Model comparisons show that accounting for mutations in essential proteins, whether through the total number $\Psi_{\text{tot_ess}}$ (ΔAICc = -0.89), the sum of PROVEAN scores $\Sigma_{\text{tot_ess}}$ (ΔAICc = -0.67), or subsetting based on the PROVEAN category (whether $\Psi_{\text{del_ess}}$ c = -1.89 or $\Psi_{\text{abs_ess}}$ ΔAICc = -1.85) slightly outperforms the null model (see Figure 3.2). The model of choice is the number of essential proteins with

mutations scored as deleterious by PROVEAN, $\Psi_{\text{del_ess}}$. These results are, however, not compelling as the ΔAICc values are all less than two in magnitude, and none of the P values is small enough to withstand any correction for multiple tests.

3.4.3 *C. reinhardtii* dataset

We began evaluating linear mixed effect models using ancestral line as a random factor. There was no predictive effect of the number of mutant proteins Ψ_{tot} (t value = 0.6, P = 0.54), the number of deleterious mutant proteins Ψ_{del} (t value = -0.5, P = 0.6), the number of mutant proteins with absolute score greater than 2.5 Ψ_{abs} (t value = -0.6, P = 0.5), or the total sum of PROVEAN scores Σ_{tot} (t value = -0.2, P = 0.8). Model comparisons show that none of the models evaluated had a lower AICc than the null model (with the only predictor being the random effect of ancestral background).

Because competitive fitness decreased significantly during MA in only two out of six ancestral backgrounds, we decided to also explore models using the absolute value of relative competitive fitness. Again, there was no predictive effect of Ψ_{tot} (t value = -1.4, P = 0.16), Ψ_{del} (t value = -0.248, P = 0.8), Ψ_{abs} (t value = -0.1, P = 0.9), or Σ_{tot} (t value = 0.7, P = 0.5). None of the models evaluated had a lower AICc than the null model (with the only predictor being the random effect of ancestral background).

Even if we separated the dataset into the different genetic backgrounds, we found only one case where a model including information from PROVEAN was the model of choice. In the 1952 background Ψ_{abs} provided a significant explanatory effect (t value = -2.99, P = 0.01), and it was chosen over the null model (which had no predictor variable) with ΔAICc of 4.3. Again, however, the effect was modest and not robust to a multiple-comparisons correction.

3.4.4 Differences in the number of clusters and supporting sequences between species

While the accuracy of PROVEAN is reported to be robust to the number of clusters used for scoring, there is a significant drop in accuracy for very small number of clusters for the human variants analyzed in the original study (Figure S4 in Choi et al., 2012). In addition, if the number of supporting sequences chosen to make out the clusters falls below 50, accuracy decreases (Figure 9 in Choi et al., 2012). 122 out of 1534 yeast proteins that we analyzed involved fewer than 30 clusters (8%), and 79 proteins had fewer than 50 supporting sequences (5%). Even more (13%) of the *C. reinhardtii* proteins received fewer than 30 clusters (187 out of 1444), and three quarters of the proteins (1082 out of 1444, 75%) had fewer than 50 supporting sequences. This lack of supporting sequences likely reflects an underrepresentation of algal proteins in the NCBI protein database, which leads to few homologous sequences passing the E value threshold of 0.1. This is evident also from the average E value score of proteins, which is an order of magnitude bigger in Cr1 than in Sc1 and Sc2 (19.9×10^{-4} compared to 1.6×10^{-4} for Sc1 and 2.9×10^{-4} for Sc2). This means that the scores for protein changes in Cr1 were based on alignment scores to few sequences with low homology, which is expected to reduce accuracy.

3.4.5 Scores for protein changes between laboratory strains and natural isolates

While we found few proteins with differences between the laboratory strain and reference genome in the two yeast datasets (142 out of 1880 proteins), the majority of proteins in Cr1 differed between the laboratory strains and the reference genome (1136 out of 1369, Table 3.2). These protein changes between MA ancestor and the reference genome most often were complex, involving more than one kind of mutation (substitution, insertion, and/or deletion). We next explored how this preexisting natural variation would be scored by PROVEAN. That is, we coded the protein variants between each laboratory strain and the reference genome and submitted these

variants with the protein of the laboratory strain as the query sequence. We have no *a priori* reason to believe that the proteins of the reference genome, on average, would be better than those of other laboratory strains. Because, as previously stated, the choice of query sequence may influence the supporting sequence set used for scoring, we also submitted variants computed with the protein of the reference genome as the query sequence. Our expectation is that scores for protein variants will be scored symmetrically around 0 (such that the absolute score will be the same regardless of whether ancestor or reference is used as the query).

For the 126 proteins that differed between the S288C reference genome and the ancestral SEYa strains used to initiate the MA experiment in Sharp et al. (2018), 103 received a neutral score (> -2.5) when the ancestral protein was mutated into the reference, and only 23 mutations were predicted to be deleterious (< -2.5). When we used the reference as the query sequence, however, we received almost twice as many deleterious scores: 43 of the ancestral variants were predicted to decrease function in comparison to the protein of the reference genome. 95 of our comparisons got exactly symmetrical scores around 0. Of the remaining 31, 21 received a score of larger magnitude when the ancestral protein was used as the query (meaning that the reference protein was better aligned to the gathered sequence clusters than the ancestors' protein).

We did the same analysis for the 14 proteins that differed between the reference genome and the ancestral diploid BY4743 strain used to initiate the MA experiment in Liu & Zhang (2019). When the ancestral protein was mutated into the reference, four of the proteins were predicted to lose function (scores < -2.5). When the reference protein was mutated into the ancestor, only one was predicted to lose function. This runs counter to the results of Sc1, where the ancestor's protein was scored as more deleterious than the reference. 11 of our comparisons got exactly symmetrical scores around 0.

Finally, in Cr1, where six different laboratory strains of *C. reinhardtii* founded by natural isolates across North America (Morgan et al. 2014) were used to initiate the MA experiment, PROVEAN scores for 2038 variants were computed. Several of these were different transcripts of the same gene. We let each gene be represented by the lowest score given to any of its transcripts. That reduces the set down to 1830 variants. The reason why there are more variants between ancestor and reference than in the MA set analyzed above is because we considered any genes that were mutated in the MA lines, but sometimes the mutation between ancestor and MA line was synonymous or occurred in introns, while there were still significant differences between the reference genome and the ancestral protein. Out of 1598 protein variants run, 90 received a score below -2.5 when submitted with the ancestral protein as the query protein. In contrast, when we used the reference genome as the query sequence 1114 of the variants received deleterious scores. According to the common interpretation of PROVEAN, this would be read as proteins of the reference genome being more functional compared to the laboratory ancestors. We compare the number of clusters and supporting sequences gathered when submitting the ancestors' protein and when submitting the reference genome protein. We find that reference proteins on average gather two more clusters of homologous sequences as compared to the ancestors' proteins. While gathering more clusters, the supporting sequence sets collected also have higher homology to the reference than to the ancestor (mean E-value of sequence set when ancestor used as query sequence = 2.6×10^{-3} , when reference used as query sequence = 1.7×10^{-3} , t-value = -4, $P < 10^{-4}$). We find differences in the absolute value of the score given in 1174 cases. Because the preexisting variants in Cr1 are dominated by complex mutations with large changes to the protein, for which PROVEAN is not evaluated, we also conducted our comparison using only proteins differing by a single amino acid substitution between reference and ancestor. Out of this subset of 114 protein

changes, 85 received perfectly symmetrical scores around 0, and only 15 received scores with absolute magnitude greater than 2.5. These deleterious mutations were equally distributed between ancestor (7 scores < -2.5 when reference was submitted as query) and reference (8 scores < -2.5 when ancestor was submitted as query). The bias in scores given to reference and ancestor mainly arises when the ancestral protein is quite different from the reference. The more different the query protein submitted is to the sequences on the NCBI database, the lower the alignment score. While PROVEAN attempts to control for differences in similarity score of query sequence to the NCBI database by using the difference in alignment score between query and mutant sequence (rather than the absolute alignment score), we believe that biases in the database likely will affect results whenever large indel mutations are considered. This also emphasizes the importance of using the correct query sequence for studies of mutational load in wild populations. One should first incorporate any shared differences to the reference into the protein before evaluating any additional polymorphism. Otherwise, natural variation risks being penalized simply by their lack of representation in the database.

3.5 Discussion

Accurately predicting the impact of mutations on fitness is a highly pursued goal of molecular population genetics. Several tools assume that beneficial alleles sweep to fixation and deleterious mutations are purged, such that neither exist as polymorphisms in standing genetic variation, while alleles with nearly neutral effects can be found at higher frequencies in standing genetic variation. Such differences in the site frequency spectrum underlie methods such as PROVEAN. The use of these tools that predict mutational effects on protein function is widespread throughout the life sciences (e.g. Table 3.1). Mutation accumulation (MA) experiments offer a way to evaluate these

methods, as the effect of mutations on organismal fitness is measured in controlled environments within similar genetic backgrounds.

In this study, we ask whether the predictions of the functional effect of variants from a frequently used tool, PROVEAN, can explain variation in growth rate in three independent sets of MA lines. In two *S. cerevisiae* datasets (Sc1 and Sc2), assigning mutations to the PROVEAN categories of deleterious and neutral does not add explanatory power. Similarly, we found no independent effect of PROVEAN in the *C. reinhardtii* (Cr1) dataset.

We believe that the most likely reason for the weak and contradictory correlations of PROVEAN scores with growth rates of MA lines may be explained by variability in how protein changes affect fitness. While PROVEAN tries to answer how mutations affect the functionality of proteins, the resulting effect on fitness will be a function of the role of the protein in the organism and the environment in which we measure fitness. For comparison, Lind, Arvidsson, Berg, & Andersson (2017) measured the competitive fitness of bacteria with mutations in single genes. They found that PROVEAN scores correlated with fitness for mutations when considering mutations in a transcription factor (AraC), enzyme (AraD), and transporter (AraE) protein. In contrast, PROVEAN scores for mutations in two ribosomal proteins (RP S20 and RP L1) did not significantly correlate with fitness. By focusing on mutations in specific genes, such studies greatly reduce the variation in fitness effects among genes. The researchers also point to the limit of their method to measure small effects on fitness, which is relevant to our study as well. In all cases, the variance in fitness among MA lines is significantly larger, but not by much, than within replicates of the same MA line. Hence, the noise in our estimates of fitness may swamp any true signal of protein functionality on fitness. This is especially true for MA datasets where lines often have many small-effect mutations spread across their genome. Adding information on the essentialness

of the protein narrows the dataset down to mutations we expect to have larger effects on fitness and should increase the correlation of fitness and PROVEAN score. However, because the number of protein changes to essential genes is low in any given line, our power to detect differences among MA lines is decreased.

Another potential reason for the low predictability of PROVEAN scores on growth rate that stem from our data is the environment in which we measure fitness. Growth rates in the lab do not necessarily have any relevance for how a certain genotype would fare in the wild. This has been most clearly demonstrated through the yeast knock-out project, where less than half of the knocked-out ORFs had a quantitative effect on growth rate in the laboratory condition used (Winzeler et al. 1999). This has consequences when analyzing the outcome of new mutations, which can have either a weaker effect than they would have under more stressful conditions or have a stronger effect than they would experience in a different genetic background. In fact, mutations scored as deleterious by PROVEAN can be adaptive in stressful conditions (Gorter et al. 2017).

Another potential caveat of our results is that not all mutations were detected. In Sc1 sites were callable at 95% of the genome. Given a total number of 1130 protein altering variants, it is possible our analyses missed some ($1130 \times 0.05 =$) 57 mutated proteins, spread out over the 218 lines. In addition, we ignored 10 mitochondrial mutations that could affect fitness. In the whole genome sequencing of the *C. reinhardtii* lines, about 75.4 Mbp of the genome was successfully genotyped (out of 120 Mbp, 62.8%). As such, we cannot exclude the possibility of large effect mutations carried in unsequenced parts of these genomes. Given a total number of 1533 protein altering variants, our analyses might miss (1533×0.372) 570 altered proteins. We also excluded

mutations in mitochondrial and chloroplast DNA in this dataset (4 cases), which may also have effects on fitness.

Finally, we did an exploratory analysis of PROVEAN score comparing the reference genome and laboratory ancestor. While no strong pattern could be seen in the scores given to pre-existing variants in the ancestors of Sc1 and Sc2, the large complex mutations that existed in the six ancestral samples in Cr1 received large negative scores when evaluated compared to the reference genome. We consider two possible explanations for this outcome. Firstly, because we only consider proteins that had mutations in the MA lines, while the reference and ancestor differ in vastly more proteins (Craig et al. 2019), it is possible that the proteins we analyze are subject to sampling bias. Perhaps previous deleterious mutations in these proteins make them more likely to accumulate mutations during MA, and as such are overrepresented in our analysis. The PROVEAN prediction would then accurately reflect the functional state of these proteins as compared to the reference genome. Alternatively, there is a bias in the sequences reported to NCBI such that sequence changes to proteins in commonly used strains and genetic model organisms compose a large portion of the database, while proteins found in less studied taxa and natural isolates are underrepresented. We report that the mean E-value (a likelihood score reflecting the homology to the query sequence) for the supporting sequence set used to compute the PROVEAN score is ten times lower when running yeast proteins compared to *Clamydomonas reinhardtii* proteins, and that there is a significant difference in the E-value when submitting the protein of the *C. reinhardtii* reference as compared to proteins of the field isolates used to initiate MA in Cr1. We encourage further studies of how the composition of the supporting sequence set (in terms of similarity score to the query protein) affect the delta score assigned by PROVEAN.

3.6 Conclusion

While bioinformatic tools that assign the vast amounts of variation found in natural populations into binary categories may be attractive in their simplicity, our results indicate that care must be taken in extrapolating PROVEAN scores to adaptive costs or benefits. The program was developed and evaluated using a database of disease-causing alleles in human populations, and not intended to quantify the fitness effects of many mutations with weak effects. We encourage further studies evaluating the applicability of mutation scoring algorithms in eco-evolutionary settings.

Chapter 4: Ploidy specific effects of mutations

4.1 Introduction

From the very first description of mutations in the form of newly arisen polyploids in plants by de Vries (1909) there have been disputes regarding the phenotypic and fitness effects of mutations (Eyre-Walker and Keightley 2007). Importantly, mutations seem to have different effects depending on the genetic background in which they appear (Chandler et al. 2014) — for sexual species with a strict alternation of ploidy, selection may act on a mutation to different degrees or potentially in opposite directions in the haploid or diploid phase.

4.1.1 Ploidy specific effects of mutations

Among the many distinguishing characteristics of life is the number of copies of an organism's genome they carry. While many organisms spend the majority of their life with one copy of their genome (most bacteria, but also multicellular haplont organisms such as dictyostelid slime moulds and green alga of the genus *Chara*), others have two (most animals), or many (polyploid plants). Yet others exist as multicellular forms in both ploidy stages, called haploid-diploid or haplodiplontic (many green, brown, and red macroalgae). Even in multicellular diplont organisms such as humans, the difference in selection in the haploid and diploid cell are of great importance. For example, mutations that render fitness advantages in gametes may have unpredicted consequences for a developing zygote (Immler 2019).

Studies measuring differences in the reproductive fitness of haploid and diploid stages of the life cycle have discovered a strong interactive effect of the environment (Gerstein & Otto, 2009; Mable, 2001; Mable & Otto, 1998; Thornber, 2006; Zörgö et al., 2013). Many of these studies are conducted in budding yeast, *Saccharomyces cerevisiae*, that can be maintained in both its haploid and diploid form. The researchers have reported broad differences in growth rate or

morphology of the two ploidy levels. Nevertheless, few studies have been conducted to measure mutational effects on fitness in the two different states.

Gerstein (2013) compared the fitness and phenotypic effect of 40 different nystatin-tolerant mutations in *S. cerevisiae* and found that haploids acquired stronger resistance to the toxin and acquired greater increases in growth rate compared to their homozygous diploid counterparts. Szafraniec, Wloch, Sliwa, Borts, & Korona (2003) measured differences in growth rate of mutant haploids created by mutagenesis and heterozygous diploids but did not report a comparison of the homozygous diploids to haploids. Here, we ask if fitness effects of spontaneous mutations from a mutation-accumulation (MA) experiment (Sharp et al. 2018) differ between ploidy levels.

4.1.2 Dominance

The genetic dominance of a mutation describes the difference in phenotype with or without the presence of a functional version. There are mutations that are imperceptible in their heterozygous state but lethal in their homozygous state and as such have a dominance of 0. In contrast, the expansion of a repeat in the huntingtin gene results in Huntington's disease regardless of the state of its fellow allele (Myers 2004), and has a dominance of 1. The average dominance coefficient will determine the likelihood of beneficial alleles to spread and fix in a population (known as Haldane's sieve, Haldane, 1927). Conversely, the frequency of deleterious mutations at mutation-selection balance will be inversely proportional to their dominance coefficient (Haldane 1937). Whether through population structure (Whitlock 2002) or inbreeding (Morton 1971), the dominance coefficient will determine the mutation load in populations with non-random mating. Understanding the average as well as the variance in dominance of spontaneous mutations is of interest in both conservation biology and medicine.

There is also reason to believe that dominance effects may correlate with mutational effect and type of protein. The shape of this relationship determines the possible mutational load and inbreeding depression of a species. While outside the scope of this study, the underlying mechanism for differences in dominance is one of the oldest and debated topics in evolutionary biology (Billiard and Castric 2011).

In addition to determining fixation probabilities, mutation frequencies, and the impacts of inbreeding, the dominance coefficient h can explain the prevalence of haploid-diploid life cycles. We should expect a diploid-dominant life cycle if deleterious mutations are sufficiently recessive in diploids (Perrot et al. 1991), while a haploid-dominant life cycle is expected when deleterious mutations are dominant (Jenkins and Kirkpatrick 1994; Jenkins and Kirkpatrick 1995). While it is commonly assumed that the mutational effect on fitness in a haploid is equal to that of a diploid homozygous for the mutation, recent theoretical work has demonstrated that intrinsic differences in how mutations affect the haploid and diploid stage of a life cycle can explain the stable co-existence of haploid-diploid organisms (Scott and Rescan 2017). By assigning separate fitness effects of mutations in the homozygous diploid (s_d) and haploid (s_h) phase, the authors identified conditions under which predominantly haploid (“haplont”), predominantly diploid (“diplont”), or biphasic life cycles (“haploid-diploid”) might evolve.

There have been several attempts to measure the distribution of dominance coefficients (Table 1 in Agrawal & Whitlock, 2011; and Table 1 in Manna, Martin, & Lenormand, 2011). Data from the yeast knockout project (Shoemaker et al. 1996) were used to estimate the distribution of dominance coefficients (Phadnis and Fry 2005; Agrawal and Whitlock 2011; Manna et al. 2012), revealing a negative correlation between selection strength and dominance. One limit of knockout data is that they are by definition, large effect mutations, with a complete loss of function in the

homozygous diploid and do not represent the multitudes of small effect mutations that are likely to arise spontaneously. As F. Manna et al. (2012) pointedly note, the yeast knockout data was not created to answer questions of dominance. In addition, these studies have not included organisms of different ploidy levels. There is a need for datasets of mutational effects measured with high replication and precision in haploids, heterozygous, and homozygous diploids to help parameterize models of life cycle evolution.

4.1.3 Our contribution

In this study, we use previously described MA lines (Sharp et al. 2018) to produce lines of different ploidy and copies of the MA genome: from haploid lines to heterozygous and homozygous diploid. Our goal was to measure how selective effects in haploids correlate with the selective effects in homozygous diploids, while at the same time measuring the average and variance of dominance effects in heterozygous yeast.

4.1.4 Justification for method

A commonly used method to produce homozygous diploids from haploids is by mating type switching. Haploid yeast exists in two different mating types, *MATa* and *MAT α* . All yeast have both mating type loci present in their genome, but one is inactive. Wild yeast are “homothallic” and can, via gene conversion using an endonuclease called HO, switch which mating type is active. This is called the “cassette model” of mating type switching, by simile to the tapes of a cassette player. Homothallism makes it possible for haploid cells in the absence of mating partners to switch the mating type they express and mate with a neighboring cell. In contrast, laboratory strains of yeast are typically heterothallic and cannot switch mating type, due to a deletion mutation of the *HO* locus. To produce homozygous yeast, researchers can insert the *HO* locus on a plasmid into the haploid yeast, which will readily switch mating type and mate. The problem with using

this approach to compare mutational effects across ploidy levels is that the diploid that is formed will have a different mating type, namely *MATa/MAT α* , and will therefore differ in more than simply the copy number of their genome (Birdsell and Wills 1996).

An alternative approach, used by Gerstein (2012), is to temporarily transform *MATa* haploids with a plasmid containing the opposite mating type. With sufficient copies of the *MAT α* plasmid the transformed haploid yeast will act like a *MAT α* haploid and mate readily with untransformed *MATa* cells. Through propagation post mating, the plasmid with the added mating type is lost, and the researcher has acquired a diploid homozygote of the *MATa/MATa* genotype. This method is not without risk however, as it can allow for further mating of the *MATa/MATa* diploids to haploids still possessing the *MAT α* -containing plasmid, leading to lines with higher ploidy levels. Furthermore, it requires first transforming all haploids of interest into *MATa* haploids by the method described above.

In this experiment, we introduce a new method. By deleting both the *MAT* locus and the *STE4* locus we render haploid yeast sterile without mating types. We then temporarily transform the yeast with two different kinds of plasmids which contain the lost genes and restore mating ability. Once mating has occurred, the diploid is propagated to lose the plasmid, and will have a *mat Δ 0/mat Δ 0* mating type. This method avoids the effect of mating type altogether.

Unfortunately given our goals, in the present study transformations induced large changes in fitness that obscure the smaller mutational effects we wanted to measure. In particular, we found high rates of loss in respiratory function, also known as petiteness (respiratory deficiency known to decrease growth rate) in yeast. Even when considering sets of yeast lines with the same petite status, we found aberrant growth patterns across the haploid, heterozygous, and homozygous states. This indicates that our experimental lines changed in more ways (genetically,

epigenetically, or otherwise) than those easily scored by petiteness and that these changes have greater impacts on growth rate than the mutations in the MA lines. We conclude that the haploid, heterozygous, and homozygous genotype of each MA line has acquired new variation, preventing us from extracting any information on the dominance of the small-effect spontaneous mutations acquired during MA. This conclusion highlights the challenge of estimating the effects of spontaneous mutations across ploidy levels: methods designed to hold “all else equal” instead induce differences.

4.2 Methods

4.2.1 Strain creation

We used 100 haploid lines of both mating types that had gone through MA (Sharp *et al.* 2018), as well as 33 replicates of their *SEY6211* ancestor (*MATa/α*, *ho*, *leu2-3 112*, *ura3-52*, *his3-Δ200*, *trp1-Δ901*, *ade2-101*, *suc2-Δ9*). Half of the lines were *rdh54Δ::KANMX*. The *RDH54* gene is involved in recombinational repair and its effect on mutation accumulation in diploids were studied in the original MA study (Sharp *et al.* 2018). The lines were streaked out from frozen on yeast peptone glycerol (YPG) agar plates to verify that the lines had functional respiratory pathways. Single colonies were then streaked out in patches on yeast peptone dextrose plates supplemented with additional adenine (YPAD), to inhibit reversion of the *ade2* mutation.

4.2.2 Transformation protocol

We grew overnight cultures from single colonies that were used to grow competent cells (cultures in exponential phase of growth). The culture of competent cells was washed with sterile water and a lithium acetate solution before we added transformation mix (containing the PCR product of interest, ssDNA, 50 % PEG, water and 1M lithium acetate). The cells were heat shocked with the

transformation mix for 1 hour at 42 degrees C. The cells were then spun down and resuspended in water before being plated on selective media corresponding to the transformation at hand.

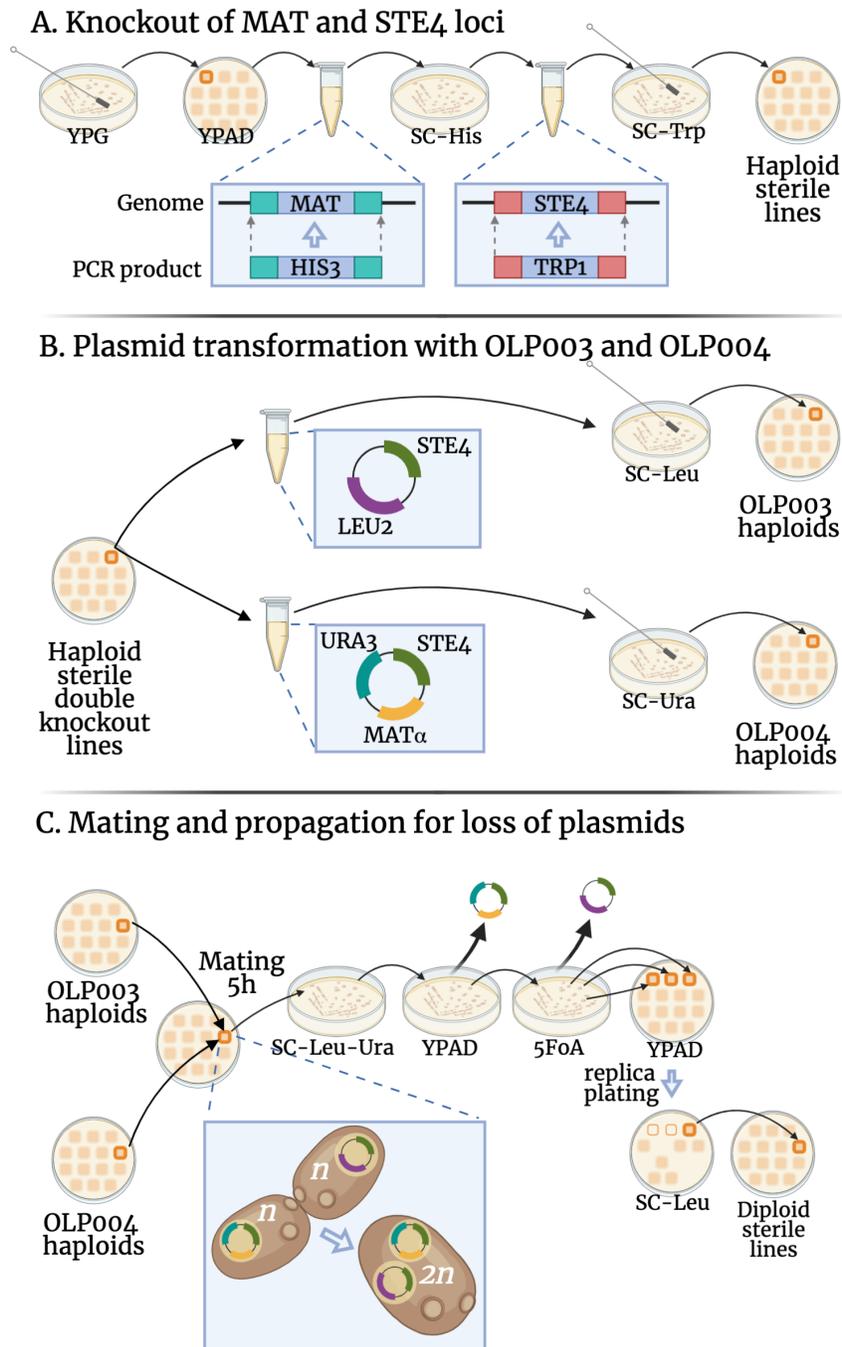


Figure 4.1 Transformations done to lines in this experiment

4.2.2.1 Step 1: Deletion of the *STE4* locus

To avoid mating during transformation of our *MAT α* lines into Δmat transformants (that act as *MAT α* and therefore could mate with their non-transformant *MAT α* siblings) we chose to make our

lines sterile by deleting the *STE4* locus. We deleted the *STE4* locus by replacing it with the *TRP1* gene (for which our original lines were deletion mutants). We amplified the *TRP1* locus from the pFA6a-*TRP1* plasmid, using primers that contain homologous sequences upstream and downstream of *STE4* and bind to regions flanking *TRP1* on the plasmid. The resulting PCR product was used to delete *STE4* with *TRP1*.

Transformation was done through 1-hour heat shock at 42 degrees C. Yeast that were successfully transformed and had their *STE4* locus replaced by *TRP1* were selected through growth on agar plates with synthetic medium lacking tryptophan (SC-Trp). A single colony from the SC-Trp plate was then streaked to single colonies again on SC-Trp. These colonies were used to verify the knockout through PCR and to proceed with strain creation.

4.2.2.2 Step 2: Deletion of the *MAT* locus

We deleted the *MAT* locus in our lines by insertion of the *HIS3* gene (for which our original lines were deletion mutants). We amplified the *HIS3* locus from the pFA6a-His3MX6 plasmid. The primers used contain homologous sequences upstream and downstream of *MAT* and bind to regions flanking *HIS3* on the plasmid. The resulting PCR product was used to delete the *MAT* locus with *HIS3*.

Transformation was done through 1-hour heat shock at 42 degrees C. Yeast that were successfully transformed and had their *MAT* locus replaced by *HIS3* were selected through growth on agar plates with synthetic medium lacking histidine (SC-His). A single colony from the SC-His plate was then streaked to single colonies again on SC-His. These colonies were used to verify the knockout through PCR. Once confirmed, a single colony from the plate was inoculated into 2 mL of YPAD media, grown at 200 rpm at 30 degrees C to saturation, and frozen in 15 % glycerol (500 μ L of saturated culture mixed with 500 μ L of 30 % glycerol).

One line (line ID 52) that had accumulated a mutation in *HIS4* in the MA experiment failed to grow on the SC-His plate post transformation and was dropped from the experiment. The 132 lines that had both the *MAT* and *STE4* locus deleted were given new identifiers as knockout (KO) lines, grown to saturation and frozen in 15 % glycerol. These lines were used in Step 3 below, as well as in dataset 1 (DS1).

4.2.2.3 Step 3: MA to KO comparison

To control for off-target insertion during the knockout of the *MAT* and *STE4* loci we ran growth rate assays for the lines before and after the knockout procedure. We grew both the original lines with their ancestors and the lines with the deletions. 20 μ L of frozen stock of these lines were inoculated in 2 mL YPAD media at 30 degrees C at 200 rpm for 2 days, after which the saturated culture was transferred to microcentrifuge tubes and stored at 4 degrees C, where they were kept for the duration of the growth rate assay. The lines were randomized across days and wells in the growth assay. To set up the growth assay, each culture was diluted 1:121 with YPAD medium. The cultures were grown for up to 24 hours, at 30 degrees C with continuous shaking set on Medium on a Bioscreen C machine. The optical density (OD) at 600 nm wavelength of the cultures was measured every 15 minutes. We conducted 11 replicate measurements of each line.

We used the spline fitting method to extract the maximum slope of the growth curve, as described by Gerstein et al. (2012). In our models, initial OD is the mean OD from 45 minutes to 2 hours (5 OD measurements). We ignore the two first OD measurements to allow the culture to reach the set temperature and be thoroughly mixed. Initial OD is taken as a proxy for the quality of the medium (a higher OD of the blank medium corresponds to more caramelization during autoclaving, leading to less accessible nutrients).

We computed the difference in mean growth rate of each line before and after knockout of *STE4* and *MAT* and flagged four samples that had a z-score below -2.5. The analysis presented exclude these samples.

We fit a linear mixed effect model of maximum slope as a function of KO (wildtype or knockout of *STE4* and *MAT*), MA (MA or control), *RDH54* status (wildtype or deletion mutant), and initial OD, to find the average effect of the *STE4* and *MAT* knockout. We included day and machine as random effects. Each line had a line identity number (line ID) from the previous mutation-accumulation experiment, which is used as a random effect, and is the same for the line before and after knockout.

The presence of genetic variance in the data is verified by fitting a simpler model (excluding KO and MA) to each of four groups: the control lines and MA lines before and after knockout. The significance of the genetic variance in explaining the data is evaluated by analysis of variance (ANOVA) of the model including line ID and a model excluding it.

4.2.2.4 Step 4: Plasmid transformation of the haploid lines

All KO lines were separately transformed with two plasmids: OLP003 contains the *STE4* and *LEU2*, OLP004 contains the *MAT α* , *STE4*, and *URA3* loci. This way, OLP003 transformant yeast will mate like *MAT α* cells and OLP004 transformant yeast will mate like *MAT α* . A single colony from each plasmid-transformed line was streaked to grow in a patch on a plate, from which we sampled the line for mating in creation of the heterozygous and homozygous versions of the lines, as well as in the creation of the haploid state in DS2 (see below).

The derivation of diploid lines was conducted twice, which we refer to as dataset 1 (DS1) and dataset 2 (DS2). In DS1, the haploid state was represented by the haploid line after knockout. In DS2, each line was randomly chosen (with two exceptions, see below) to have either the

OLP003 transformant or their OLP004 transformant represent their haploid state. We sampled the chosen patch with a toothpick and streaked it out to single colony on YPAD twice, after which three colonies were patched on both YPAD and leucine (in the case of OLP003 transformant) or uracil (for OLP004 transformant) drop-out plates. The patch that grew on YPAD but not leucine/uracil drop-out plates was inoculated in YPAD, grown for two days in a 30 degrees C shaking incubator and then frozen into aliquots with 15 % glycerol.

One line (line ID 39) failed to be transformed with OLP004. We chose the OLP003 transformant to represent this line's haploid state. One line (line ID 83) with a preexisting mutation in *LEU2* failed to grow on the SC-Leu plate post transformation with OLP003. We suspect *LEU2* inserted during the *MAT* locus deletion was insufficient to rescue this mutant. We chose the OLP004 transformant to represent this line's haploid state. One line (line ID 84) could not be successfully transformed with either plasmid and was dropped from the experiment.

4.2.2.5 Step 5: Mating

To create the heterozygote state, a swab from the OLP003 transformant was patched on a plate together with a random OLP004 transformant from one of the control lines. For the homozygote state, the swab from the OLP003 transformant was patched on a plate together with its OLP004 transformant. The lines were allowed to mate for 5 hours before being transferred and spread out with a cotton tip on double drop-out -uracil -leucine plates. After two days of growth a colony growing on the double drop-out was picked and streaked out on YPAD. After another two days of growth a single colony was chosen and streaked out on an agar plate containing 5-Fluororotic acid (5FoA), on which only yeast with a deficient uracil pathway can grow, to confirm loss of the OLP004 plasmid. After three days of growth, three colonies were taken and patched on YPAD and leucine drop-out, to confirm loss of the OLP003 plasmid. After two days, the patch that grew on

YPAD but not on leucine drop-out was inoculated in YPAD, grown for two days in a 30 degrees C shaking incubator and then frozen into aliquots with 15 % glycerol.

4.2.2.6 Step 6: Fitness and ploidy assays

The growth rates of DS1 were measured together with the double knockout haploid lines prior to their transformation in August 2019. The growth rates of DS2 were measured together with haploids after their plasmid-transformation in January 2020. We conducted the growth rate assays as described in Step 3: MA to KO comparison.

We confirmed the ploidy of our lines by flow cytometry. Samples were inoculated from fridge cultures (15 μ L into 1 mL of YPAD with added 50 μ g mL⁻¹ of ampicillin in 96-well boxes). After one night of growth 5 μ L of each culture was transferred to 96 well assay plates and washed in water before being fixed with 70 % ethanol. The samples in ethanol were either kept at room temperature for over one hour or overnight in the fridge before proceeding. Plates were then centrifuged at 2500 RPM for 5 minutes, washed in 50 mM sodium citrate solution, and then incubated at 37 degrees C in a 50 mM sodium citrate solution with 6.25 μ g of RNAase A added to each sample. The following day the samples were spun down, supernatant removed, and the cells stained in a solution of sodium citrate and 7.5 μ L of 50 mM sytox green per sample.

The stained samples were run on an Attune NxT flow cytometer, which allows processing of 96-well plates. We ran the samples at 25 μ L min⁻¹ until 10 000 events had been measured. The resulting files were exported and analyzed with the flowCore package in R. All lines had the expected ploidy level.

4.3 Results

Data files and scripts used for analyses are available for download at <https://zoology.ubc.ca/~sandell/ploidominance/>.

4.3.1 Effect of *MAT* and *STE4* knockout

Four MA lines fell 2.5 standard deviations below the average knockout effect, with large decreases in their growth rates post knockout (line IDs 2, 106, 122 and 127). We exclude these lines in the comparison of lines before and after knockout. The knockout of *MAT* and *STE4* led to increases in growth rates of both control and MA lines (t-value 9.7, χ^2 of model with compared to without effect of KO = 79.59, $p < 10^{-5}$), see Figure 4.2A.

We found significant difference in growth rate among MA lines, but not control lines, before knockout ($\chi^2 = 42.3$, $P < 10^{-5}$ for MA lines, $\chi^2 = 3.2$, $P = 0.073$ for control lines). After knockout of the *STE4* and *MAT* loci, we verified this result ($\chi^2 = 195$, $P < 10^{-5}$ for MA lines, $\chi^2 = 1.9$, $P = 0.16$ for control lines).

Despite the promising variance in growth rates among MA lines we found no correlation between the growth rate of lines before and after knockout (Kendall's rank correlation $\tau = 0.087$, $P = 0.22$, see Figure 4.2B). This is problematic because it indicates that whatever signal existed before about the impact of mutations on growth has been swamped by variation induced by knockout.

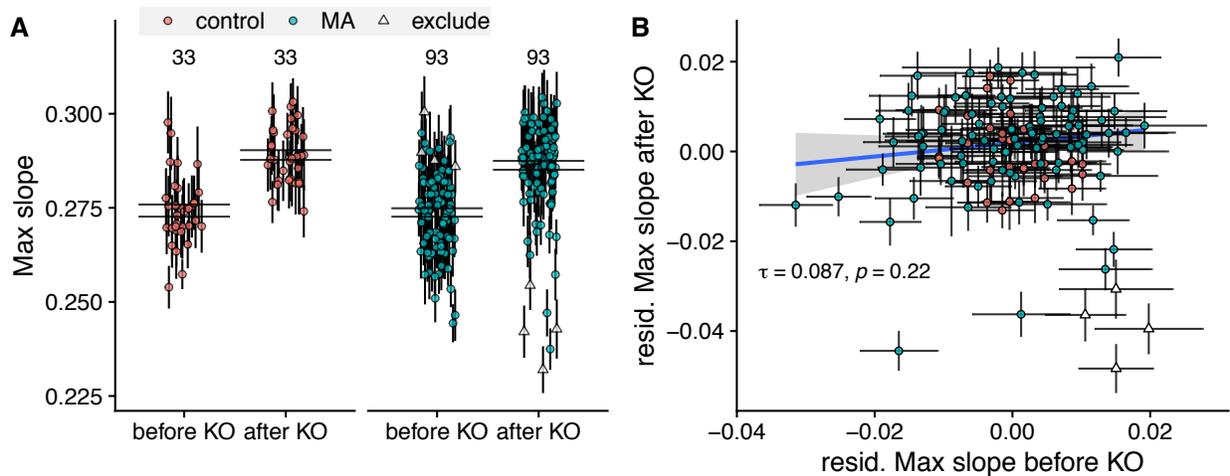


Figure 4.2 Comparison of growth rates before and after knockout of the *STE4* and *MAT* loci A) Maximum growth rates of control and MA lines increase after knockout of the *STE4* and *MAT* loci. Large black lines represent the standard error of the mean in each group. Numbers above plot show number of lines in each group. B) Correlation of growth rates (max slope) of MA lines before and after knockout. Triangles signify the four MA lines which change in growth rate fell -2.5 standard deviations below the mean and were excluded from the analysis. Colored points with grey error bars represent line means and standard error of the mean.

4.3.2 High frequency of petites following plasmid transformation

In an initial experiment (DS1), we generated diploids by combining two plasmid-transformed haploids, followed immediately by selection for plasmid loss. We found a large difference in the growth rates between haploid and diploid lines.

To determine if plasmid transformation was the source of the large decreased fitness in the diploids, we re-transformed our lines, saving the haploid lines after transformation to compare the growth of these to the diploids created from them (DS2). On average, we found that the haploids had also decreased in growth rate compared to DS1. However, we also found very strong bimodality in our data, see Figure 4.3.

To determine whether variation among lines might have been caused by loss of mitochondrial function we spotted our lines on YPG to test for petitesness. While the number of petites was low among haploids in DS1 (7 out of the 131 lines without plasmid transformation:

four of these were the four lines we detected as outliers in our comparison of growth rates before and after knockout), we found a high prevalence in the diploids (169 out of 260 lines). In DS2, where haploids were isolated post plasmid transformation, we found about an equal proportion of petites in the haploid and diploid lines (about half of our lines were petite in both groups). After further analysis, we found that diploid lines that had the *RDH54* gene intact were more likely to become petite (χ^2 for DS1 = 36, $P < 0.0001$, χ^2 for DS2 = 61, $P < 0.0001$).

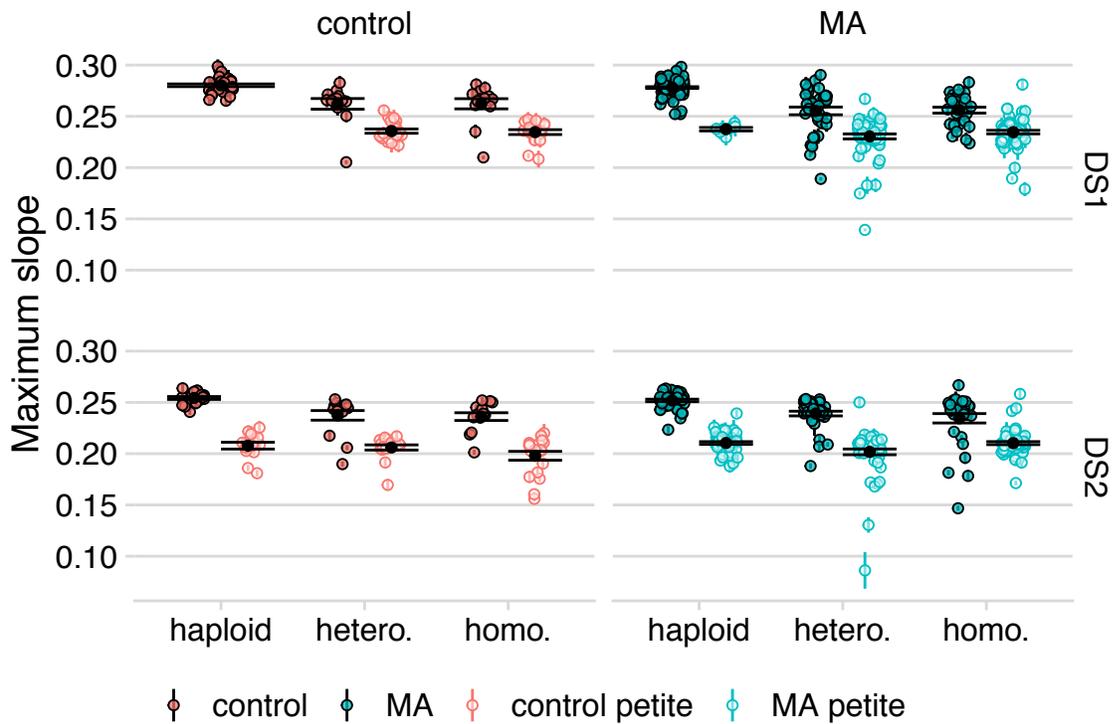


Figure 4.3 Petiteness of the line strongly predicts maximum slope of growth curve in both datasets. Large black points and lines represent the standard error of the mean in each group. Colored points with error bars represent line means and standard error of the mean.

4.3.3 No correlation of growth rates across data sets

We set out to test the correlation in growth rates among lines that presented the same petite status in both datasets. We found no significant correlation between the growth rate of lines in DS1 and

DS2 with the same genotype and petite status (Table 4.1). Hence, we analyzed the two datasets separately.

Table 4.1 Kendall's correlation coefficient of mean growth rate of MA lines with same petite status in DS1 and DS2

Genotype	Petite status	Correlation coef	P value
Haploid	Grande	0.137	0.188
Heterozygote	Petite	-0.137	0.193
	Grande	-0.029	0.890
Homozygote	Petite	0.001	0.994
	Grande	-0.0833	0.690

4.3.4 Inferences of mutational effects across haploid, heterozygote, and homozygote yeast

The linear mixed effect models we fit to DS1 and DS2 have the same structure: maximum slope is fitted as a function of petite status, *RDH54* status, MA status (whether MA or control), genotype, and initial OD as fixed effects. We also include an interaction between MA status and genotype. We control for day, machine (in DS1 only, as all DS2 replicates were measured on one machine), and haplotype effect (an identifier shared by the haploid, heterozygote, and homozygote from the same haploid genome) by including them as random effects in our model.

Neither MA nor the interaction of MA and genotype is significant in DS1 (t-value = -1.45, $p = 0.15$ for MA). Haploid lines had a significantly higher growth rate than diploids in DS1 (t-value = -14.163 and -13.950 for heterozygotes and homozygotes respectively, $p = 10^{-5}$ for both). In DS2 MA is also not significant, but the interaction between MA and homozygosity is (t-value = 5.78, $p = 10^{-5}$). This would mean that homozygous MA lines have a higher growth rate than homozygous control lines.

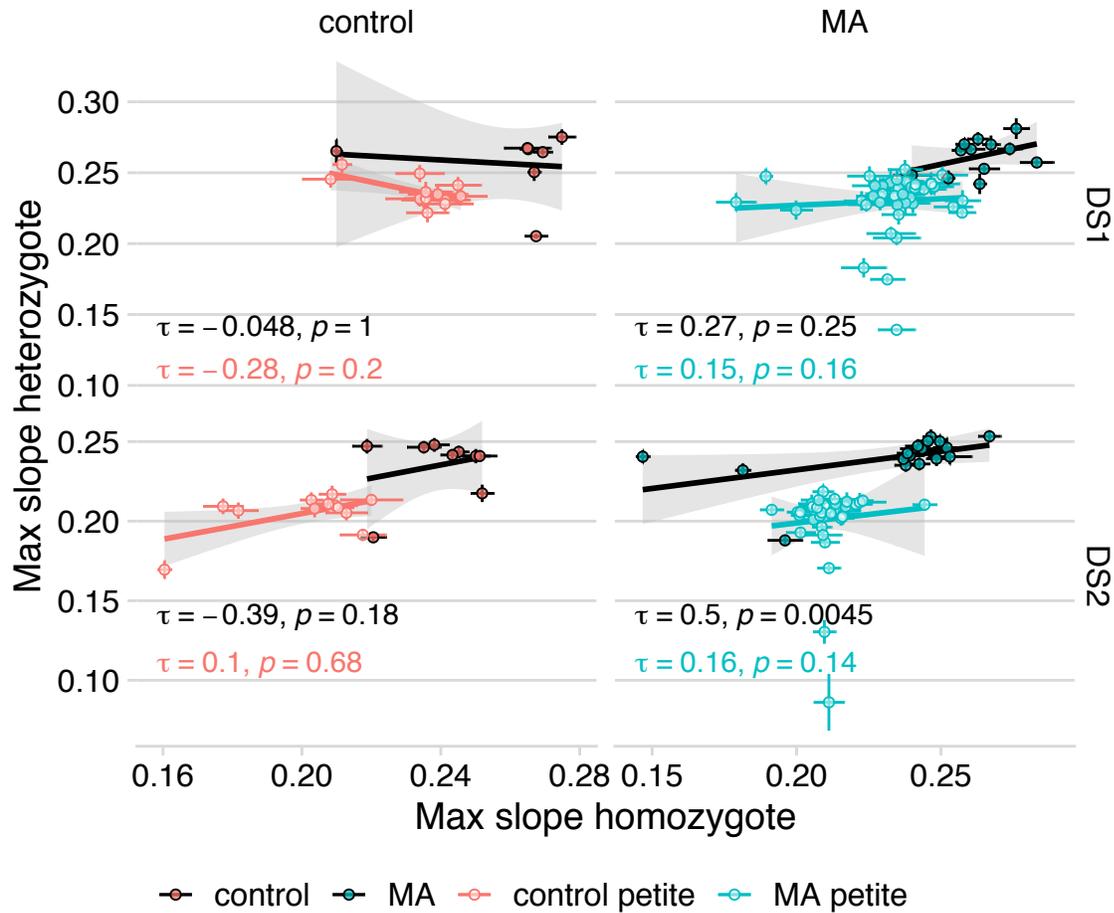


Figure 4.4 We find no correlation between the growth rate of the MA line in its heterozygote and homozygote form. Large black points and lines represent the standard error of the mean in each group. Colored points with error bars represent line means and standard error of the mean.

Because we found no correlation between how diploid lines of the same petite status grew in DS1 and DS2, we considered the possibility that unique fitness-affecting mutations occurred in each line during transformation. In effect, this would mean that the aim of our study, to compare the growth rate of identical mutated genomes in three different genotypic conditions, would be compromised. If this were the case, we would not expect a correlation between the fitnesses of the haploid and diploid homozygous state of each haplotype, because the fitness effects of pre-existing mutations would be swamped by new and different mutations in the two states. We compared our

original model that used MA identity as a random effect on intercept (with the same identity for the haploid and homozygous lines derived from the same MA line) to a model using line identity as a random effect (that treats each genotype as its own group). In this latter model, there is no information on the original haplotype used to establish the line. The model using line identity rather than haplotype is preferred in both DS1 ($\Delta\text{AIC} = -351$) and DS2 ($\Delta\text{AIC} = -1,286$), suggesting that our MA mutations are overwhelmed by unique fitness-affecting mutations that occurred in each line during transformation. These results remained unchanged when excluding the four lines that were outliers in the KO to MA comparison.

4.4 Discussion

The fitness effects of spontaneous mutations depend on the genotypic state of the organism. While it is often recognized that a deleterious mutation will not be fully recessive (*i.e.* heterozygote mutants will differ from wildtype), less attention has been paid to differences in the fitness effect of mutations across ploidy levels (*i.e.* haploid mutants may differ in fitness from homozygote mutant diploids). Previous studies have been limited to single genes or considered only one aspect of the question (comparing only heterozygous and homozygous mutants, or only haploid and diploid mutants). In this project, we set out to conduct a large-scale experiment to measure both dominance within diploids and differences between ploidy levels using 100 previously established haplotypes with fitness-affecting spontaneous mutations in yeast. By establishing haploid, heterozygote, and homozygote genotypes of each haplotype, we aimed to study the fitness effect of mutations in each of these genotypic states with enough replicated fitness assays that the fitness of each line could be accurately measured. In addition, we disrupted the mating pathway in all of our yeast, to control for differential expression between mating types.

To establish the three genotypic states of our mutant genomes (bearing MA haplotypes) we put our lines through a number of transformations. These transformations led to high rates of loss of respiratory function (petites), accompanied with low growth rates. The combination of heat shock (Van Uden 1985) and serial bottlenecks (Taylor et al. 2002) potentially account for the high frequency of petites. There is evidence that the *his3-Δ200* allele carried by our MA ancestor increases the formation of petites under heat stress (Zhang et al. 2003). Usually the loss of respiratory function is an easily distinguished mutant phenotype and the genetic underpinnings can be numerous (Dimitrov et al. 2009). The petite lines in our experiment were not of the small colony variety, which is why they were not originally caught. Even if the petite mutation was the same across our experimental replicates, the fitness interaction between mitochondrial and nucleic mutations is often significant (Zeyl et al. 2005a).

Our ability to measure fitness effects in the lab is limited. The fitness effect of the spontaneous mutations accumulated through 100 bottlenecks were small and both deleterious and beneficial (Sharp et al. 2018). By contrast, the changes in fitness that arose during the genetic transformations were of much larger effects. Heat shock and consecutive bottlenecks likely allowed genetic or epigenetic changes to fix in our experimental replicates.

This is a cautionary tale. Genetic transformations are commonplace in experimental studies of microbes. We believe that our study highlights the importance both of designing experiments such that treatment groups mimic each other as close as possible, to control for unforeseen effects of laboratory manipulation. Indeed, it seemed counterintuitive to us to isolate and use the haploids after plasmid transformation (rather than before). Our initial analysis of DS1, in which haploids were isolated prior to plasmid transformation, showed large differences in the growth rate of haploid and diploid lines. As our results of DS2 show, in which haploids were isolated post-

transformation, this effect was driven largely by detrimental effect of undergoing plasmid transformation. By attempting to hold “all else” constant, the method we used introduced variance greater than what we set out to measure. We designed our study with the explicit purpose of measuring small effects, which made the fitness effect of the transformation easy to spot. Even so, it required a close to zero growth rate of a particular line for us to consider the possibility of induced defects. Even within the subset of lines that did not have respiratory defects, we did not find a signal of the mutations we set out to measure. This can be both due to additional mutations arising in the formation of the haploid, diploid heterozygous, and diploid homozygous lines and a limit to the precision of our method. Indeed, even with competition experiments assayed with flowcytometric methods Gallet, Cooper, Elena, & Lenormand (2012) found unexpected variability in their mutational fitness estimates, which they attribute to cryptic variation.

Measuring small mutational effects across ploidy and genotype levels presents a major challenge. A previous attempt to measure the dominance effects of mutations found that heterozygous mutants became homozygous during the fitness assay, literally changing as they were being measured (Gerstein et al. 2014). Similarly, in our experiment, yeast lines changed while we transformed them into the desired genotype and ploidy.

4.4.1 Deletion of *RDH54* protects against petite-formation

We found an unexpected pattern where the *rdh54* Δ deletion mutants were less likely than the wildtype to become petite following transformations. Petite yeast are generally more heat tolerant than wildtype yeast (Van Uden 1985). *rdh54* Δ deletion mutants at stationary phase have also been shown to be tolerant to heat shock (Jarolim et al. 2013). Even though our strains were subjected to heat shock for transformation in the exponential phase, there may be interactions between temperature, the Rdh54 protein, and mitochondrial function that accounts for the pattern.

A plethora of genes are involved in mitochondrial genome stability (see *Wide-scale screening for rho0 production* in Contamine & Picard, 2000). Although there is no evidence as far as we know that the Rdh54 protein is involved in mitochondrial repair or recombination (Chen 2013), the suggestion has been made that petite yeast could arise from homologous recombination between imperfect duplicates (Faye et al., 1973; Gaillard, Strauss, & Bernardi, 1980; Slonimski & Lazowska, 1977; Weiller, Bruckner, Kim, Pratje, & Schweyen, 1991; see also the “Recombination/repair track” in Contamine & Picard, 2000), and Rdh54 is a vital protein for homologous recombination in the nucleus. Thus another explanation is that the deletion of *RDH54* could reduce mitochondrial recombination rates.

4.5 Conclusion

Knowledge of the selection strength of spontaneous mutations in haploid, heterozygous diploids, and homozygous diploids is relevant for the evolution of all sexually reproducing organisms with alternation of ploidy levels. To measure the selective coefficient of small effect mutations we need to generate a large number of mutated genomes in all three genotypic states. Our study shows that the common methods used for transforming yeast into the desired type (ploidy or genotype) led to large changes in fitness that cloud the signal of the mutations of interest and thus make comparisons across ploidy and genotype challenging.

Chapter 5: Conclusion

Yeast make an excellent study organism. Thanks to the malleability of its life cycle, we are able to study evolutionary processes from genomic evolution to shifts in mating systems. I want to highlight the two studies that most inspired my thesis work: McDonald et al. (2016) and Sharp et al. (2018). I will describe their methods and findings and how my research has built and added to their findings.

5.1 Sex, drugs, and experimental evolution

In the first year of my doctoral studies McDonald et al. (2016) published a landmark study comparing the molecular evolution in sexual and asexual populations of yeast. The results, that sex allows yeast to remove deleterious mutations while maintaining beneficial ones and as such evolve higher growth rates than populations that are restricted to clonal reproduction, confirmed theoretical predictions from decades (Fisher 1930; Muller 1932). I had studied mathematical models of the evolution of sex during my master's program in Europe and had sought to work with Dr. Otto because of her excellent work on the topic. The graphs showing how beneficial mutations were able to sweep to fixation as deleterious mutations were recombined away in the sexual populations, while asexual populations were stuck dragging deleterious mutations along, as the theory of selective interference had predicted (Crow and Kimura 1965; Maynard Smith 1971; Weissman and Barton 2012) convinced me of the power of experimental evolution.

The yeast strains that were used in the McDonald et al. (2016) study were genetically modified in a number of ways. The two mating types carried different antibiotic markers to make them distinguishable from each other and the diploid that would result from mating the two. The yeast also carried auxotrophic markers under ploidy and mating-type specific promoters. This made it possible to select against potential diploids by germinating and growing the haploid

cultures using nutrient-limited medium in which only cells of the right mating type and ploidy could grow. It seemed to me then a perfectly engineered system, and the opportunities endless.

For my first project I wanted to insert two new antibiotic markers, close to the original two, and alternate the combination of markers that would be favored in each sexual cycle, to simulate a case of rapidly fluctuating selection that has been shown to favor an increase in recombination (Maynard Smith 1971). In short, I wanted to evolve longer map length between two genetic markers. This had been done previously in *Drosophila melanogaster* (Charlesworth and Charlesworth 1985a,b; Charlesworth et al. 1985), but with the sequence technology available today we would be able to find the genetic basis for any observed increase in recombination. While we were able to insert the new markers in the genome in the location that we wanted, the markers did not segregate in the expected ratios. Rather, it seemed that the genetic manipulations caused problems during meiosis, because I saw a drastic decrease in spore viability in the newly created lines as compared to the original McDonald et al. (2016) lines and unexpected marker combinations. I spent a year troubleshooting my methods and gathering the data necessary to convince myself that the intended experiment could not be done.

At about the same time that I put the final nail in the coffin of my first project, my colleague Dr. Nathaniel Sharp was about to give up on his own experiment using the McDonald et al. (2016) lines. His idea was to alter the ratio of the two mating types during mating, to alter the strength of competition for mating opportunities. He had, however, found a rapid decrease in sporulation rate during the first sexual cycles, rendering further sexual cycles impossible. I needed a new project, and my acquaintance with the strains and methods made me confident that I could find a solution. The main difference between the experimental life cycle used in the McDonald et al. (2016) study and the cycle used in my chapter 2 (apart from the skew in mating ratios) was that I combined

auxotrophic and antibiotic selection during germination. The initial issues with decreased sporulation rates found by Dr. Sharp were still a worry to me, and I started to consider what other aspects of the life cycle would respond to the frequent rounds of meiosis we would put the yeast through.

I forced yeast to go through alternating rounds of asexual and sexual reproduction under different asymmetric mating ratios. The lines where *MAT α* cells had to compete more strongly for mating opportunities evolved to become more efficient maters than the lines that experienced other mating ratios. Hence, there is an asymmetry between the two mating types in response to sexual selection. My hypothesis is that this result arises from physiological differences between the *MAT α* and *MAT a* cells. While the enzyme that controls the mating response in *MAT α* cells (by degrading the *MAT a* pheromone) is cell-associated, its counterpart in *MAT a* cells is released into the medium. Banderas et al. (2016) proposed that this fact could help *MAT a* cells assess the relative ratio of *MAT a* to *MAT α* cells (and hence their opportunity for mating). I believe that the increased mating efficiency observed in Chapter 2 could be achieved by mutations to these enzymes. Reducing the degradation of the opposite mating-type's pheromone will increase the propensity to mate. Because this degradation occurs in the extracellular environment in *MAT a* cells, individual mutations to this enzyme are unlikely to change the overall concentration of *MAT α* pheromone. Thus while isogamous at a morphological level, the mating types of *S. cerevisiae* are differentiated at molecular and functional levels.

I hope that the second chapter of my thesis will spur further experimental studies into the mating and meiosis machinery of yeast. Foundational and current papers showing increased rates of adaptation in sexually reproducing yeast cultures (Zeyl and Bell 1997; Goddard et al. 2001, 2005; Gray and Goddard 2012; McDonald et al. 2016; Kosheleva and Desai 2017; Leu et al. 2020)

often report only the rate of clonal growth in the asexual and sexual cultures. Other adaptations to a more sexual life cycle remain to be revealed. Separately from the comparisons of evolution in sexual and asexual cultures, there has been research on the mating process in yeast (Rogers and Greig 2009; Smith and Greig 2010; Smith et al. 2013). The one study that included mating competition during evolution found that mating competition decreased adaptive rates such that the benefit of sexual reproduction was lost (Reding et al. 2013). Though some data on the mating rate of these evolved populations were obtained they did no comparisons of the response to sexual selection in the two different mating types. My study shows that the response is markedly different depending on which mating type is put under strong competition for mating. I believe that further studies should be done using many different strains of yeast to find out if this result is replicable.

The idea of a sexual rather than asexual control in the experiment was also Dr. Sharp's. By having a control line that goes through sex without experiencing selection for mating opportunities we could disentangle the effects of mating competition from the effects of sexual reproduction. For me, one of the most striking results of my second chapter was the convergent evolution of haploids that were heterozygous at the mating type locus, subverting the sexual control that I had designed (lines with plenty of mating opportunities). The underlying cause for this evolution is still a mystery to me, and I would like to understand when and how these haploids spread. Was it survival in the diploid selective medium post-germination that gave them an edge? Was there a relative growth advantage of these haploids over diploids during clonal growth that increased their frequency? Or was it the ability to survive the digestive enzyme zymolyase post sporulation, without the need to go through the costly process of meiosis, that drove them to evolutionary victory? Further experiments are needed to answer these questions. Studies using different strains and selective markers to see if this result is replicable would also be valuable.

5.2 Fitness effect of mutations

I was lucky to start my PhD during Dr. Sharp's time in the Otto lab. He invited me to contribute to a mutation accumulation (MA) experiment in haploid and diploid yeast in the second year of my PhD. Other MA experiments in yeast had either used diploid (Nishant et al. 2010; Zhu et al. 2014) or haploid strains (Lynch et al. 2008; Serero et al. 2014). The results from these studies suggested that haploids may experience higher single nucleotide mutation (SNM) rates than diploids, but there had been no formal comparison between the mutational rates of the two ploidy states.

Ours was the first MA study to directly measure and compare mutation accumulation in haploids and diploids. One possible explanation for a higher SNM rate in haploids could be that diploids repair mutations through recombinational repair with the second chromosome copy. To test this hypothesis, we deleted the *RDH54* gene in the haploid and diploid ancestor in half of the lines. This gene is involved in recombinational repair during mitosis in diploids, but not in haploids (Klein 1997). We propagated 200 lines (about 50 lines of each ploidy and *RDH54* status) by single colony transfers for 100 days (~ 336 000 generations). We found that the rate of mitochondrial mutations (both single nucleotide and insertion-deletion) and single nucleotide mutations overall were higher in haploid lines compared to diploid lines (Sharp et al. 2018), but we found no difference between diploids with or without the *RDH54* gene. Indeed, the only significant effect of the *RDH54* deletion was an increased rate of chromosomal gain and loss in diploids.

Because I had studied the effects of mutational bias on the evolution of sex during my master program (Vanhoenacker et al. 2018), I was most interested in the phenotypic and fitness effects of the mutations that were accumulated. Strikingly, we found a decrease in growth rate of our diploid MA lines compared to the ancestor, but no difference in mean growth rates of our

haploid lines before and after MA (Sharp et al. 2018). This was unexpected since diploids are supposed to be able to mask deleterious mutations with their extra copy of each gene. We would have expected haploids to show greater loss in fitness than the diploids. We found that the main driver of the diploid loss in growth rate was aneuploidy, gains of new copies of chromosomes. I was eager to find out the relationship between mutational effects on fitness in haploid and diploid yeast when all else was equal. Even though we had found no mean difference in growth rates of haploids before and after MA, we found genetic variation in fitness in the haploid MA lines that was not present in the control lines. The fitness effects of these mutations became the focus of my third and fourth chapter.

My third chapter lays in the intersection of two fields, and I found myself oscillating between two questions. One was how to map genotype to fitness in MA lines, and the other was whether bioinformatic tools could aid us in this pursuit. Mapping genotype to fitness in MA lines poses many challenges (Eyre-Walker and Keightley 2007), one being that MA lines have many different kinds of mutations scattered throughout the genome, and so ascribing fitness declines to specific mutations is difficult. We had found that chromosomal abnormalities explained a substantial portion of the decreased growth in the diploid MA lines, although the number of nonsynonymous mutations also had a significant negative effect on fitness in diploid, but not in haploid, lines (Sharp et al. 2018). Again, it was Dr. Sharp who aired the idea that PROVEAN predictions for the functional effect of mutations may help us distinguish between different kinds of nonsynonymous mutations and help us draw the genotype to fitness map in our MA lines. He also suggested using a second MA dataset in *C. reinhardtii* where the fitness and mutations of lines were known (Morgan et al. 2014; Ness et al. 2015; Kraemer et al. 2017).

Kraemer et al. (2017) had found that differences between *C. reinhardtii* MA lines and controls were well explained by the number of coding sequence mutations. They also observed that the best model for describing the distribution of fitness effects of mutations was a binary categorization for each mutation as either neutral or deleterious to fitness. Sequence alignment-based approaches, such as PROVEAN, are restricted to nonsynonymous coding sequence mutations (*i.e.* amino acid changes in proteins) and deliver functional predictions in binaries: ‘deleterious’ or ‘neutral’ in PROVEAN (Choi et al. 2012) and ‘intolerant’ and ‘tolerant’ in SIFT (Ng and Henikoff 2001). The fact that I was not able to find a significant explanatory effect of the number of mutations scored as ‘deleterious’ by PROVEAN either in our own MA lines (Sharp et al. 2018), a different set of MA lines in yeast (Liu and Zhang 2019), or in the *C. reinhardtii* lines (Ness et al. 2015; Kraemer et al. 2017) tells us that the designation into the two categories of fitness-affecting mutations do not correspond clearly to functional changes in the proteins (Chapter 3).

The study by Lind et al. (2017) gives some insights into why the PROVEAN predictions failed to explain changes in fitness. In that study different types of protein changing mutations were induced in a set of genes in the bacterium *Salmonella typhimurium*. The effect of each mutation on growth rate was tested in isolation, and the PROVEAN prediction compared to the measured effect on fitness. Five different proteins were mutated: a transcription factor, an enzyme, a transporter, and two ribosomal proteins. They found a significant correlation between PROVEAN prediction and fitness effect of mutations in the transcription factor, the enzyme, and transporter, but not in the ribosomal proteins (Lind et al. 2016). The authors ascribe these differences to different mutational robustness of the different kinds of proteins. Our MA lines naturally have mutations in many different kinds of proteins, only some of which contribute

substantially to fitness (less than half of the genes in the yeast genome have a measurable effect on fitness when deleted, Winzeler et al. 1999). It is not just the functional effect of a mutation on the protein that matters when we draw our genotype to fitness map, but also the relative importance of each protein to the overall phenotype. We can partly control for this in our yeast lines, where we know the essentialness of proteins, and we show that adding this information to our model improves the correlation between mutations and growth rates. This result echoes the metabolic control theory of dominance (Kacser and Burns 1981) which I described in the introduction, in that only enzymes with large control coefficients on the flux will substantially alter the phenotype.

In the fourth chapter of my thesis, we set out to compare the mutational effects of fitness in haploids and diploids. We used the haploid MA and control lines from our previous study (Sharp et al. 2018) to create heterozygous and homozygous diploids carrying the same mutations. The first aim of this study was to look for correlations in the fitness of the haploid and homozygous diploids lines. The second was to look for correlations in fitness between the heterozygous diploid and the haploids and homozygotes, providing estimates of dominance.

Previous research in the Otto lab had shown that mutations for antibiotic resistance affected fitness differently in haploid and diploid yeast. Gerstein (2012) compared the growth rate of haploid and homozygous diploid lines carrying the same adaptive mutations and found that haploid yeast on average had larger fitness gains relative to the haploid control than did the homozygous diploid to its control. Still, the assumption persists that the fitness effects of deleterious mutations in the haploid and homozygous diploid state are equal, as exemplified by the use of haploid rather than homozygous diploid lines to measure dominance coefficients of gene deletions in yeast (Marek and Korona 2016).

We found no correlation between the growth rate of haploid and homozygous diploids in either of our two attempts at conducting this experiment. We also found no correlation between the growth rate of the heterozygous and homozygous diploids carrying the same mutation. We also found no correlation between how the same diploid line grew the first and second time we created them. As we found that about half of our lines in each category (haploid, heterozygous diploid, homozygous diploid) had acquired mutations to their respiratory function, rendering them petite (unable to respire) with varying effects on growth, as well as potentially other mutations, I concluded that we could no longer answer my original questions.

Our result mirrors that of a previous Otto lab member, Dr. Aleeza Gerstein, who attempted to measure dominance of beneficial adaptive mutations. During the growth assays, her heterozygous lines became homozygous, replicating the beneficial mutation into both chromosome copies (Gerstein et al. 2014). This loss of heterozygosity was only caught because of the large fitness effect of the mutations on antibiotic resistance, which Dr. Gerstein studied. Similarly, if it was not for the drastic reduction in fitness of the lines that acquired petite mutations in Chapter 4, I would not have thought to question the genetic changes in the haploid versus homozygous versus heterozygous diploid lines carrying the same original mutations. In general, these results add to the difficulties in measuring the dominance of small effect mutations that others have previously outlined (Gallet et al. 2012; Manna et al. 2012).

There is one result that stood out clearly in the fourth chapter, namely the strong association between status of the *RDH54* gene and the propensity for petite mutations to arise. The underlying causes for this remain unknown to us, and I'm excited to see molecular biologists pursue research on this relationship.

In conclusion, my doctoral research has discovered that the evolutionary response to sexual selection is markedly different between the two different mating types of yeast, which adds to the literature on the diversity and evolution of reproductive systems. I have also shown that functional effects on the level of the protein (as estimated by PROVEAN) do not necessarily correlate with the mutational effect on fitness, which I believe should be taken into consideration especially in eco-evolutionary studies that search for adaptive loci or estimate mutational load. Finally, my fourth chapter illustrates how the amenability of the yeast genome that experimental biologists' treasure in fact has a treacherous side: our experimental subject changes before our eyes. It should never be presumed that "all else is equal" when it comes to yeast.

References

- Agrawal, A. F., and M. C. Whitlock. 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187:553–566.
- Al-Shuhaib, M. B. S., F. R. Al-Kafajy, M. A. Badi, S. AbdulAzeez, K. Marimuthu, H. A. I. Al-Juhaishi, and J. F. Borgio. 2018. Highly deleterious variations in COX1, CYTB, SCG5, FK2, PRL and PGF genes are the potential adaptation of the immigrated African ostrich population. *Comput. Biol. Med.* 100:17–26.
- Al Khatib, H. A., A. A. Al Thani, and H. M. Yassine. 2018. Evolution and dynamics of the pandemic H1N1 influenza hemagglutinin protein from 2009 to 2017. *Arch. Virol.* 163:3035–3049.
- Alavijeh, E. S., J. Khajehali, S. Snoeck, R. Panteleri, M. Ghadamyari, W. Jonckheere, S. Bajda, C. Saalwaechter, S. Geibel, V. Douris, J. Vontas, T. Van Leeuwen, and W. Dermauw. 2020. Molecular and genetic analysis of resistance to METI-I acaricides in Iranian populations of the citrus red mite *Panonychus citri*. *Pestic. Biochem. Physiol.* 164:73–84.
- Alfano, N., V. Tagliapietra, F. Rosso, U. Ziegler, D. Arnoldi, and A. Rizzoli. 2020. Tick-borne encephalitis foci in northeast Italy revealed by combined virus detection in ticks, serosurvey on goats and human cases. *Emerg. Microbes Infect.* 9:474–484.
- Alvarez, J. B., L. Castellano, R. Recio, and A. Cabrera. 2019. Wx Gene in *Hordeum chilense*: Chromosomal Location and Characterisation of the Allelic Variation in the Two Main Ecotypes of the Species. *Agronomy* 9:261.
- Bajda, S., W. Dermauw, R. Panteleri, N. Sugimoto, V. Douris, L. Tirry, M. Osakabe, J. Vontas, and T. Van Leeuwen. 2017. A mutation in the PSST homologue of complex I (NADH:ubiquinone oxidoreductase) from *Tetranychus urticae* is associated with resistance to METI acaricides. *Insect Biochem. Mol. Biol.* 80:79–90.
- Banderas, A., M. Koltai, A. Anders, V. Sourjik, B. Alvaro, M. and Koltai, and A. Alexander, and S. Victor. 2016. Sensory input attenuation allows predictive sexual response in yeast. *Nat. Commun.* 7:12590.
- Bazhenov, M., A. Chernook, P. Kroupin, G. Karlov, and M. Divashuk. 2020. Molecular Characterization of the Dwarf53 Gene Homolog in *Dasyphyrum Villosum*. *Plants* 9:186.
- Ben-Ari, G., D. Zenvirth, A. Sherman, L. David, M. Klutstein, U. Lavi, J. Hillel, and G. Simchen. 2006. Four Linked Genes Participate in Controlling Sporulation Efficiency in Budding Yeast. *PLOS Genet.* 2:1–9.
- Bhattacharya, S., S. Dhar, A. Banerjee, and S. Ray. 2019. Structural, functional, and evolutionary analysis of late embryogenesis abundant proteins (LEA) in *Triticum aestivum*: A detailed molecular level biochemistry using in silico approach. *Comput. Biol. Chem.* 82:9–24.
- Billiard, S., and V. Castric. 2011. Evidence for Fisher’s dominance theory: How many “special cases”? *Trends Genet.* 27:441–445. Elsevier Ltd.
- Birdsell, J., and C. Wills. 1996. Significant competitive advantage conferred by meiosis and syngamy in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 93:908–912.
- Böndel, K. B., S. A. Kraemer, T. Samuels, D. McClean, J. Lachapelle, R. W. Ness, N. Colegrave, and P. D. Keightley. 2019. Inferring the distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii*. *PLoS Biol.*, doi: 10.1371/journal.pbio.3000192.
- Buhl, M., C. Kästle, A. Geyer, I. B. Autenrieth, S. Peter, and M. Willmann. 2019. Molecular Evolution of Extensively Drug-Resistant (XDR) *Pseudomonas aeruginosa* Strains From Patients and Hospital Environment in a Prolonged Outbreak. *Front. Microbiol.* 10.
- Bulmer, M. G., and G. A. Parker. 2002. The evolution of anisogamy: a game-theoretic approach. *Proc. R. Soc. London. Ser. B Biol. Sci.* 269:2381–2388.
- Cabrera, A., L. Castellano, R. Recio, and J. B. Alvarez. 2019. Chromosomal location and molecular characterization of three grain hardness genes in *Agropyron cristatum*. *Euphytica* 215:165.

- Cargill, M., D. Altshuler, J. Ireland, P. Sklar, K. Ardlie, N. Patil, C. R. Lane, E. P. Lim, N. Kalyanaraman, J. Nemesh, L. Ziaugra, L. Friedland, A. Rolfe, J. Warrington, R. Lipshutz, G. Q. Daley, and E. S. Lander. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.*, doi: 10.1038/10290.
- Castellá, G., M. R. Bragulat, R. A. Cigliano, and F. J. Cabañes. 2020. Transcriptome analysis of non-ochratoxigenic *Aspergillus carbonarius* strains and interactions between some black aspergilli species. *Int. J. Food Microbiol.* 317:108498.
- Chandler, C. H., S. Chari, D. Tack, and I. Dworkin. 2014. Causes and consequences of genetic background effects illuminated by integrative genomic analysis. *Genetics*, doi: 10.1534/genetics.113.159426.
- Charlesworth, B. 1994. Evolutionary Genetics: The nature and origin of mating types. *Curr. Biol.* 4:739–741.
- Charlesworth, B., and D. Charlesworth. 1985a. Genetic variation in recombination in *Drosophila*. I. Responses to selection and preliminary genetic analysis. *Heredity (Edinb)*. 54:71–83.
- Charlesworth, B., and D. Charlesworth. 1985b. Genetic variation in recombination in *Drosophila*. II. Genetic analysis of a high recombination stock. *Heredity (Edinb)*. 54:85–98.
- Charlesworth, B., I. Mori, and D. Charlesworth. 1985. Genetic variation in recombination in *drosophila*. III. Regional effects on crossing over and effects on non-disjunction. *Heredity (Edinb)*. 55:209–221.
- Chasnov, J. R. 2000. Mutation-selection balance, dominance and the maintenance of sex. *Genetics* 156:1419–1425.
- Chen, X. J. 2013. Mechanism of Homologous Recombination and Implications for Aging-Related Deletions in Mitochondrial DNA. *Microbiol. Mol. Biol. Rev.* 77:476–496.
- Cheng, P., C. R. Gedling, G. Patil, T. D. Vuong, J. G. Shannon, A. E. Dorrance, and H. T. Nguyen. 2017. Genetic mapping and haplotype analysis of a locus for quantitative resistance to *Fusarium graminearum* in soybean accession PI 567516C. *Theor. Appl. Genet.* 130:999–1010.
- Choi, Y., and A. P. Chan. 2015. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, doi: 10.1093/bioinformatics/btv195.
- Choi, Y., G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan. 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One*, doi: 10.1371/journal.pone.0046688.
- Codón, A. C., J. M. Gasent-Ramírez, and T. Benítez. 1995. Factors which affect the frequency of sporulation and tetrad formation in *Saccharomyces cerevisiae* bakers yeasts. *Appl. Environ. Microbiol.* 61:630–638.
- Contamine, V., and M. Picard. 2000. Maintenance and Integrity of the Mitochondrial Genome: a Plethora of Nuclear Genes in the Budding Yeast. *Microbiol. Mol. Biol. Rev.* 64:281–315.
- Conte, G. L., K. A. Hodgins, S. Yeaman, J. C. Degner, S. N. Aitken, L. H. Rieseberg, and M. C. Whitlock. 2017. Bioinformatically predicted deleterious mutations reveal complementation in the interior spruce hybrid complex. *BMC Genomics* 18:970.
- Craig, R. J., K. B. Böndel, K. Arakawa, T. Nakada, T. Ito, G. Bell, N. Colegrave, P. D. Keightley, and R. W. Ness. 2019. Patterns of population structure and complex haplotype sharing among field isolates of the green alga *Chlamydomonas reinhardtii*. *Mol. Ecol.* 28:3977–3993.
- Crow, J. F., and M. Kimura. 1965. Evolution in Sexual and Asexual Populations. *Am. Nat.*, doi: 10.1086/282389.
- Danaher, R. J., D. E. Fouts, A. P. Chan, Y. Choi, J. DePew, J. M. McCarrison, K. E. Nelson, C. Wang, and C. S. Miller. 2017. HSV-1 clinical isolates with unique in vivo and in vitro phenotypes and insight into genomic differences. *J. Neurovirol.* 23:171–185.
- de Lomana, A. L. G., A. Kaur, S. Turkarslan, K. D. Beer, F. D. Mast, J. J. Smith, J. D. Aitchison, and N. S. Baliga. 2017. Adaptive Prediction Emerges Over Short Evolutionary Time Scales. *Genome Biol. Evol.* 9:1616–1623.
- de Vries, H. 1909. The mutation theory; experiments and observations on the origin of species in the vegetable kingdom. Open Court Pub. Co., Chicago.

- del Olmo, I., L. Poza-Viejo, M. Piñeiro, J. A. Jarillo, and P. Crevillén. 2019. High ambient temperature leads to reduced FT expression and delayed flowering in *Brassica rapa* via a mechanism associated with H2A.Z dynamics. *Plant J.* 100:343–356.
- Deutschbauer, A. M., and R. W. Davis. 2005. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* 37:1333–1340.
- Deutschbauer, A. M., R. M. Williams, A. M. Chu, and R. W. Davis. 2002. Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* 99:15530–15535.
- Diaz Caballero, J., S. T. Clark, P. W. Wang, S. L. Donaldson, B. Coburn, D. E. Tullis, Y. C. W. Yau, V. J. Waters, D. M. Hwang, and D. S. Guttman. 2018. A genome-wide association analysis reveals a potential role for recombination in the evolution of antimicrobial resistance in *Burkholderia multivorans*. *PLoS Pathog.* 14:e1007453.
- Dimitrov, L. N., R. B. Brem, L. Kruglyak, and D. E. Gottschling. 2009. Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains. *Genetics* 183:365–383.
- Eddy, S. R. 2004. Where did the BLOSUM62 alignment score matrix come from? *Nat. Biotechnol.* 22:1035–1036.
- Ellis, B., P. Haaland, F. Hahne, N. Le Meur, N. Gopalakrishnan, J. Spidlen, M. Jiang, and G. Finak. 2019. flowCore: flowCore: Basic structures for flow cytometry data.
- Enyenihi, A. H., and W. S. Saunders. 2003. Large-Scale Functional Genomic Analysis of Sporulation and Meiosis in *Saccharomyces cerevisiae*. *Genetics* 163:47–54.
- Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8:610–618.
- Faye, G., H. Fukuhara, C. Grandchamp, J. Lazowska, F. Michel, J. Casey, G. S. Getz, J. Locker, M. Rabinowitz, M. Bolotin-Fukuhara, D. Coen, J. Deutsch, B. Dujon, P. Netter, and P. P. Slonimski. 1973. Mitochondrial nucleic acids in the peptite colonic mutants: Deletions and repetitions of genes. *Biochimie* 55:779–792.
- Ferchaud, A.-L., M. Laporte, C. Perrier, and L. Bernatchez. 2018. Impact of supplementation on deleterious mutation distribution in an exploited salmonid. *Evol. Appl.* 11:1053–1065.
- Fisher, K. J., S. Kryazhimskiy, and G. I. Lang. 2019. Detecting genetic interactions using parallel evolution in experimental populations. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 374:20180237.
- Fisher, R. 1930. Chapter 6: Sexual reproduction and sexual selection.
- Fisher, R. A. 1931. The evolution of dominance. *Biol. Rev.*, doi: 10.1111/j.1469-185X.1931.tb01030.x.
- Fisher, R. A. 1928. The Possible Modification of the Response of the Wild Type to Recurrent Mutations. *Am. Nat.*, doi: 10.1086/280193.
- Gaillard, C., F. Strauss, and G. Bernardi. 1980. Excision sequences in the mitochondrial genome of yeast. *Nature* 283:218–220.
- Gallet, R., T. F. Cooper, S. F. Elena, and T. Lenormand. 2012. Measuring selection coefficients below 10^{-3} : Method, Questions, and Prospects. *Genetics* 190:175–186.
- Gammerdinger, W. J., M. A. Conte, B. A. Sandkam, D. J. Penman, and T. D. Kocher. 2019. Characterization of sex chromosomes in three deeply diverged species of Pseudocrenilabrinae (Teleostei: Cichlidae). *Hydrobiologia* 832:397–408. Springer International Publishing.
- Gerke, J. P., C. T. L. Chen, and B. A. Cohen. 2006. Natural Isolates of *Saccharomyces cerevisiae* Display Complex Genetic Variation in Sporulation Efficiency. *Genetics* 174:985–997.
- Gerstein, A. C. 2012. Mutational effects depend on ploidy level: all else is not equal. *Biol. Lett.* 9:20120614.
- Gerstein, A. C., and J. Berman. 2015. Shift and adapt: the costs and benefits of karyotype variations. *Curr. Opin. Microbiol.* 26:130–136.

- Gerstein, A. C., H.-J. E. Chun, A. Grant, and S. P. Otto. 2006. Genomic Convergence toward Diploidy in *Saccharomyces cerevisiae*. *PLOS Genet.* 2:1–6.
- Gerstein, A. C., A. Kuzmin, and S. P. Otto. 2014. Loss-of-heterozygosity facilitates passage through haldane’s sieve for *saccharomyces cerevisiae* undergoing adaptation. *Nat. Commun.* 5:1–9. Nature Publishing Group.
- Gerstein, A. C., and S. P. Otto. 2009. Ploidy and the causes of genomic evolution. *J. Hered.* 100:571–581.
- Goddard, M. R., H. C. Godfray, and B. Austin. 2005. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* 434:636–640.
- Goddard, M. R., D. Greig, and A. Burt. 2001. Outcrossed sex allows a selfish gene to invade yeast populations. *Proc. R. Soc. London. Ser. B Biol. Sci.* 268:2537–2542.
- Goodenough, U., H. Lin, and J. H. Lee. 2007. Sex determination in *Chlamydomonas*. *Semin. Cell Dev. Biol.* 18:350–361.
- Gorter, F. A., M. F. L. Derks, J. van den Heuvel, M. G. M. Aarts, B. J. Zwaan, D. de Ridder, and J. A. G. M. de Visser. 2017. Genomics of Adaptation Depends on the Rate of Environmental Change in Experimental Yeast Populations. *Mol. Biol. Evol.* 34:2613–2626.
- Gray, J. C., and M. R. Goddard. 2012. Sex enhances adaptation by unlinking beneficial from detrimental mutations in experimental yeast populations. *BMC Evol. Biol.* 12:43.
- Haber, J. E. 2012. Mating-Type Genes and MAT Switching in *Saccharomyces cerevisiae*. *Genetics* 191:33–64.
- Hadjivasiliou, Z., and A. Pomiankowski. 2016. Gamete signalling underlies the evolution of mating types and their number. *Philos. Trans. R. Soc. B Biol. Sci.* 371.
- Haldane, J. B. S. 1927. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Math. Proc. Cambridge Philos. Soc.* 23:838–844. Cambridge University Press.
- Haldane, J. B. S. 1937. The Effect of Variation on Fitness. 71:337–349.
- Hall, D. W., and S. B. Joseph. 2010. A High Frequency of Beneficial Mutations Across Multiple Fitness Components in *Saccharomyces cerevisiae*. *Genetics* 185:1397–1409.
- Hall, J. P. J., E. Harrison, and M. A. Brockhurst. 2018. Competitive species interactions constrain abiotic adaptation in a bacterial soil community. *Evol. Lett.* 2:580–589.
- Halligan, D. L., and P. D. Keightley. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.*, doi: 10.1146/annurev.ecolsys.39.110707.173437.
- Hamabata, T., G. Kinoshita, K. Kurita, P.-L. Cao, M. Ito, J. Murata, Y. Komaki, Y. Isagi, and T. Makino. 2019. Endangered island endemic plants have vulnerable genomes. *Commun. Biol.* 2:244.
- Harris, J. K., S. T. Kelley, G. B. Spiegelman, and N. R. Pace. 2003. The genetic core of the universal ancestor. *Genome Res.*, doi: 10.1101/gr.652803.
- Harrison, E., D. Guymer, A. J. Spiers, S. Paterson, and M. A. Brockhurst. 2015. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr. Biol.* 25:2034–9.
- He, C., Y. Wei, Y. Zhu, Y. Xia, D. M. Irwin, and Y. Liu. 2019. Adaptive Evolution of C-Type Lysozyme in Vampire Bats. *J. Mol. Evol.* 87:309–316.
- Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.89.22.10915.
- Henikoff, S., and J. G. Henikoff. 1991. Automated assembly of protein blocks for database searching. *Nucleic Acids Res.* 19:6565–6572.
- Henson, L., N. Songsasen, W. Waddell, K. Wolf, L. Emmons, S. Gonzalez, E. Freeman, and J. Maldonado. 2017. Characterization of genetic variation and basis of inflammatory bowel disease in the Toll-like receptor 5 gene of the red wolf and the maned wolf. *Endanger. Species Res.* 32:135–144.

- Heuermann, M. C., M. G. Rosso, M. Mascher, R. Brandt, H. Tschiersch, L. Altschmied, and T. Altmann. 2019. Combining next-generation sequencing and progeny testing for rapid identification of induced recessive and dominant mutations in maize M2 individuals. *Plant J.* 100:851–862.
- Hodgins, K. A., D. G. Bock, M. A. Hahn, S. M. Heredia, K. G. Turner, and L. H. Rieseberg. 2015. Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. *Mol. Ecol.* 24:2226–2240.
- Huberman, L. B., and A. W. Murray. 2013. Genetically Engineered Transvestites Reveal Novel Mating Genes in Budding Yeast. *Genetics* 195:1277–1290.
- Immler, S. 2019. Haploid Selection in “diploid” Organisms. *Annu. Rev. Ecol. Evol. Syst.* 50:219–236.
- Jarolim, S., A. Ayer, B. Pillay, A. C. Gee, A. Phrakaysone, G. G. Perrone, M. Breitenbach, and I. W. Dawes. 2013. *Saccharomyces cerevisiae* genes involved in survival of heat shock. *G3 Genes, Genomes, Genet.* 3:2321–2333.
- Jenkins, C. D., and M. Kirkpatrick. 1995. Deleterious mutation and the evolution of genetic life cycles. *Evolution (N. Y.)*, doi: 10.1111/j.1558-5646.1995.tb02283.x.
- Jenkins, C. D., and M. Kirkpatrick. 1994. Deleterious mutation and ecological selection in the evolution of life cycles. *Lect. Math. Life Sci.*
- Kacser, H., and J. A. Burns. 1981. The molecular basis of dominance. *Genetics* 97:639–666.
- Kadian, K., S. Vijay, Y. Gupta, R. Rawal, J. Singh, A. Anvikar, V. Pande, and A. Sharma. 2018. Structural modeling identifies *Plasmodium vivax* 4-diphosphocytidyl-2C-methyl-d-erythritol kinase (IspE) as a plausible new antimalarial drug target. *Parasitol. Int.* 67:375–385.
- Kaltz, O., and G. Bell. 2002. The ecology and genetics of fitness in *Chlamydomonas*. XII. Repeated sexual episodes increase rates of adaptation to novel environments. *Evolution (N. Y.)*. 56:1743–1753.
- Kardos, M., H. R. Taylor, H. Ellegren, G. Luikart, and F. W. Allendorf. 2016. Genomics advances the study of inbreeding depression in the wild. *Evol. Appl.* 9:1205–1218.
- Khalid, M., Z. Khalid, A. Gul, R. Amir, and Z. Ahmad. 2018. Characterization of wheat cell wall invertase genes associated with drought tolerance in synthetic-derived wheat. *Int. J. Agric. Biol.* 20:2677–2684.
- Klein, H. L. 1997. RDH54, a RAD54 homologue in *Saccharomyces cerevisiae*, is required for mitotic diploid-specific recombination and repair and for meiosis. *Genetics*.
- Klimushina, M. V., P. Y. Kroupin, M. S. Bazhenov, G. I. Karlov, and M. G. Divashuk. 2020. Waxy Gene-Orthologs in Wheat × *Thinopyrum* Amphidiploids. *Agronomy* 10:963.
- Knop, M. 2006. Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. *BioEssays* 28:696–708.
- Kosheleva, K., and M. M. Desai. 2017. Recombination Alters the Dynamics of Adaptation on Standing Variation in Laboratory Yeast Populations. *Mol. Biol. Evol.* 35:180–201.
- Kraemer, S. A., K. B. Böndel, R. W. Ness, P. D. Keightley, and N. Colegrave. 2017. Fitness change in relation to mutation number in spontaneous mutation accumulation lines of *Chlamydomonas reinhardtii*. *Evolution (N. Y.)*, doi: 10.1111/evo.13360.
- Kusakabe, M., A. Ishikawa, M. Ravinet, K. Yoshida, T. Makino, A. Toyoda, A. Fujiyama, and J. Kitano. 2017. Genetic basis for variation in salinity tolerance between stickleback ecotypes. *Mol. Ecol.* 26:304–319.
- Lachapelle, J., and G. Bell. 2012. Evolutionary Rescue Of Sexual And Asexual Populations In A Deteriorating Environment. *Evolution (N. Y.)*. 66:3508–3518.
- Lang, G. I., A. W. Murray, and D. Botstein. 2009. The cost of gene expression underlies a fitness trade-off in yeast. *Proc. Natl. Acad. Sci.* 106:5755–5760.
- Lecová, L., P. Tůmová, and E. Nohýnková. 2019. Clone-based haplotyping of *Giardia intestinalis* assemblage B human isolates. *Parasitol. Res.* 118:355–361.

- Lenth, R. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means.
- Lessells, C. M., R. R. Snook, and D. J. Hosken. 2009. The evolutionary origin and maintenance of sperm: Selection for a small, motile gamete mating type. First Edit. Elsevier Ltd.
- Leu, J. Y., S. L. Chang, J. C. Chao, L. C. Woods, and M. J. McDonald. 2020. Sex alters molecular evolution in diploid experimental populations of *S. cerevisiae*. *Nat. Ecol. Evol.* 4:453–460. Springer US.
- Li, C., Y. Zhang, J. Li, L. Kong, H. Hu, H. Pan, L. Xu, Y. Deng, Q. Li, L. Jin, H. Yu, Y. Chen, B. Liu, L. Yang, S. Liu, Y. Zhang, Y. Lang, J. Xia, W. He, Q. Shi, S. Subramanian, C. D. Millar, S. Meader, C. M. Rands, M. K. Fujita, M. J. Greenwold, T. A. Castoe, D. D. Pollock, W. Gu, K. Nam, H. Ellegren, S. Y. W. Ho, D. W. Burt, C. P. Ponting, E. D. Jarvis, M. T. P. Gilbert, H. Yang, J. Wang, D. M. Lambert, J. Wang, and G. Zhang. 2014. Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. *Gigascience* 3:27.
- Lind, P. A., L. Arvidsson, O. G. Berg, and D. I. Andersson. 2016. Variation in Mutational Robustness between Different Proteins and the Predictability of Fitness Effects. *Mol. Biol. Evol.* msw239.
- Liu, H., and J. Zhang. 2019. Yeast Spontaneous Mutation Rate and Spectrum Vary with Environment. *Curr. Biol.* 29:1584-1591.e3. Elsevier Ltd.
- Liu, Q., Y. Zhou, P. L. Morrell, and B. S. Gaut. 2017. Deleterious variants in Asian rice and the potential cost of domestication. *Mol. Biol. Evol.* 34:msw296.
- Longo, L. G. A., V. S. de Sousa, G. B. Kraychete, L. H. Justo-da-Silva, J. A. Rocha, S. V. Superti, R. R. Bonelli, I. S. Martins, and B. M. Moreira. 2019. Colistin resistance emerges in pandrug-resistant *Klebsiella pneumoniae* epidemic clones in Rio de Janeiro, Brazil. *Int. J. Antimicrob. Agents* 54:579–586.
- Lynch, M., W. Sung, K. Morris, N. Coffey, C. R. Landry, E. B. Dopman, W. J. Dickinson, K. Okamoto, S. Kulkarni, D. L. Hartl, and W. K. Thomas. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.0803466105.
- Mable, B. K. 2001. Ploidy evolution in the yeast *Saccharomyces cerevisiae*: A test of the nutrient limitation hypothesis. *J. Evol. Biol.* 14:157–170.
- Mable, B. K., and S. P. Otto. 1998. The evolution of life cycles with haploid and diploid phases. *BioEssays* 20:453–462.
- Magwene, P. M., Ö. Kayçç, J. A. Granek, J. M. Reininga, Z. Scholl, and D. Murray. 2011. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* 108:1987–1992.
- Makino, T., C.-J. Rubin, M. Carneiro, E. Axelsson, L. Andersson, and M. T. Webster. 2018. Elevated Proportions of Deleterious Genetic Variation in Domestic Animals and Plants. *Genome Biol. Evol.* 10:276–290.
- Manna, F., R. Gallet, G. Martin, and T. Lenormand. 2012. The high-throughput yeast deletion fitness data and the theories of dominance. *J. Evol. Biol.* 25:892–903.
- Manna, F., G. Martin, and T. Lenormand. 2011. Fitness landscapes: An alternative theory for the dominance of mutation. *Genetics* 189:923–937.
- Marcus, S., C. B. Xue, F. Naider, and J. M. Becker. 1991. Degradation of a-factor by a *Saccharomyces cerevisiae* alpha-mating-type-specific endopeptidase: evidence for a role in recovery of cells from G1 arrest. *Mol. Cell. Biol.* 11:1030–1039.
- Marek, A., and R. Korona. 2016. Strong dominance of functional alleles over gene deletions in both intensely growing and deeply starved yeast cells. *J. Evol. Biol.* 29:1836–1845.
- Martino, M. E., P. Joncour, R. Leenay, H. Gervais, M. Shah, S. Hughes, B. Gillet, C. Beisel, and F. Leulier. 2018. Bacterial Adaptation to the Host's Diet Is a Key Evolutionary Force Shaping *Drosophila*-*Lactobacillus* Symbiosis. *Cell Host Microbe* 24:109-119.e6.
- Maynard Smith, J. 1971. What use is sex? *J. Theor. Biol.*, doi: 10.1016/0022-5193(71)90058-0.

- McClure, A. W., K. C. Jacobs, T. R. Zyla, and D. J. Lew. 2018. Mating in wild yeast: delayed interest in sex after spore germination. *Mol. Biol. Cell* 29:3119–3127.
- McDonald, M. J., D. P. Rice, and M. M. Desai. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* 531:233–236.
- Mendel, G. J. 1865. Versuche über Pflanzen-Hybriden [Experiments in Plant Hybridization]. *Verhandlungen des naturforschenden Vereines Brünn* [Proceedings Nat. Hist. Soc. Brünn].
- Mercatanti, A., S. Lodovichi, T. Cervelli, and A. Galli. 2017. CRIMEtoYHU: a new web tool to develop yeast-based functional assays for characterizing cancer-associated missense variants. *FEMS Yeast Res.* 17.
- Morales, H. E., A. Pavlova, L. Joseph, and P. Sunnucks. 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol. Ecol.* 24:2820–37.
- Morgan, A. D., R. W. Ness, P. D. Keightley, and N. Colegrave. 2014. Spontaneous mutation accumulation in multiple strains of the green alga, *Chlamydomonas reinhardtii*. *Evolution (N. Y.)*. 68:2589–2602.
- Morton, N. E. 1971. Population genetics and disease control. *Soc. Biol.* 18:243–251. Routledge.
- Muller, H. J. 1932. Some Genetic Aspects of Sex. *Am. Nat.*, doi: 10.1086/280418.
- Myers, R. H. 2004. Huntington's Disease Genetics. *NeuroRx* 1:255–262.
- Navarro-Sigüenza, A. G., H. Vázquez-Miranda, G. Hernández-Alonso, E. A. García-Trejo, and L. A. Sánchez-González. 2017. Complex biogeographic scenarios revealed in the diversification of the largest woodpecker radiation in the New World. *Mol. Phylogenet. Evol.* 112:53–67.
- Ness, R. W., A. D. Morgan, N. Colegrave, and P. D. Keightley. 2012. Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*. *Genetics*, doi: 10.1534/genetics.112.145078.
- Ness, R. W., A. D. Morgan, R. B. Vasanthakrishnan, N. Colegrave, and P. D. Keightley. 2015. Extensive de novo mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. *Genome Res.* 25:1739–1749.
- Ng, P. C., and S. Henikoff. 2001. Predicting deleterious amino acid substitutions. *Genome Res.* 11:863–874.
- Ng, P. C., and S. Henikoff. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* 7:61–80.
- Nishant, K. T., W. Wei, E. Mancera, J. L. Argueso, A. Schlattl, N. Delhomme, X. Ma, C. D. Bustamante, J. O. Korbel, Z. Gu, L. M. Steinmetz, and E. Alani. 2010. The Baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLoS Genet.*, doi: 10.1371/journal.pgen.1001109.
- Ochoa, A., D. P. Onorato, R. R. Fitak, M. E. Roelke-Parker, and M. Culver. 2017. Evolutionary and Functional Mitogenomics Associated With the Genetic Restoration of the Florida Panther. *J. Hered.* 108:449–455.
- Ofori-Anyinam, B., A. J. Riley, T. Jobarteh, E. Gitteh, B. Sarr, T. I. Faal-Jawara, L. Rigouts, M. Senghore, A. Kehinde, N. Onyejebu, M. Antonio, B. C. de Jong, F. Gehre, and C. J. Meehan. 2020. Comparative genomics shows differences in the electron transport and carbon metabolic pathways of *Mycobacterium africanum* relative to *Mycobacterium tuberculosis* and suggests an adaptation to low oxygen tension. *Tuberculosis* 120:101899.
- Orr, H. A., and S. P. Otto. 1994. Does diploidy increase the rate of adaptation? *Genetics*.
- Otto, S. P. 2003. The advantages of segregation and the evolution of sex. *Genetics* 164:1099–1118.
- Otto, S. P. 1994. The role of deleterious and beneficial mutations in the evolution of ploidy levels. *Lect. Math. Life Sci.*
- Parker, G. A. 2011. The origin and maintenance of two sexes (anisogamy), and their gamete sizes by gamete competition. P. *in* T. Togashi and P. A. Cox, eds. *The Evolution of Anisogamy: A Fundamental Phenomenon Underlying Sexual Selection*. Cambridge University Press, Cambridge UK.
- Parker, G. A., R. R. Baker, and V. G. F. Smith. 1972. The origin and evolution of gamete dimorphism and the male-

- female phenomenon. *J. Theor. Biol.* 36:529–553.
- Parker, T. A., J. C. Berny Mier Y Teran, A. Palkovic, J. Jernstedt, and P. Gepts. 2020. Pod indehiscence is a domestication and aridity resilience trait in common bean. *New Phytol.* 225:558–570.
- Perrier, C., A.-L. Ferchaud, P. Sirois, I. Thibault, and L. Bernatchez. 2017. Do genetic drift and accumulation of deleterious mutations preclude adaptation? Empirical investigation using RADseq in a northern lacustrine fish. *Mol. Ecol.* 26:6317–6335.
- Perrot, V., S. Richerd, and M. Valéro. 1991. Transition from haploidy to diploidy. *Nature* 351:315–317.
- Peter, J., M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre, K. Freil, A. Llored, C. Cruaud, K. Labadie, J.-M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemainque, P. Wincker, G. Liti, and J. Schacherer. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556:339–344.
- Phadnis, N., and J. D. Fry. 2005. Widespread correlations between dominance and homozygous effects of mutations: Implications for theories of dominance. *Genetics* 171:385–392.
- Piombo, E., P. Bosio, A. Acquadro, P. Abbruscato, and D. Spadaro. 2020. Different Phenotypes, Similar Genomes: Three Newly Sequenced *Fusarium fujikuroi* Strains Induce Different Symptoms in Rice Depending on Temperature. *Phytopathology*® 110:656–665.
- Priyam, M., M. Tripathy, U. Rai, and S. M. Ghorai. 2018. Divergence of protein sensing (TLR 4, 5) and nucleic acid sensing (TLR 3, 7) within the reptilian lineage. *Mol. Phylogenet. Evol.* 119:210–224.
- Randerson, J. R., and L. D. Hurst. 2001. The uncertain evolution of the sexes. *Trends Ecol. Evol.* 16:571–579.
- Rasal, K. D., V. Chakrapani, S. K. Patra, S. Jena, S. D. Mohapatra, S. Nayak, J. K. Sundaray, P. Jayasankar, and H. K. Barman. 2016. Identification and prediction of the consequences of nonsynonymous SNPs in glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene of zebrafish *Danio rerio*. *Turkish J. Biol.* 40:43–54.
- Reding, L. P., J. P. Swaddle, and H. A. Murphy. 2013. Sexual selection hinders adaptation in experimental populations of yeast. *Biol. Lett.* 9:20121202.
- Renaut, S., T. Replansky, A. Heppleston, and G. Bell. 2006. the Ecology and Genetics of Fitness in *Chlamydomonas*. Xiii. Fitness of Long-Term Sexual and Asexual Populations in Benign Environments. *Evolution (N. Y.)* 60:2272.
- Renaut, S., and L. H. Rieseberg. 2015. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other compositae crops. *Mol. Biol. Evol.* 32:2273–2283.
- Rogers, D. W., and D. Greig. 2009. Experimental evolution of a sexually selected display in yeast. *Proc. R. Soc. B* 276:543–549.
- Rowland, S. D., K. Zumstein, H. Nakayama, Z. Cheng, A. M. Flores, D. H. Chitwood, J. N. Maloof, and N. R. Sinha. 2020. Leaf shape is a predictor of fruit quality and cultivar performance in tomato. *New Phytol.* 226:851–865.
- Ruderfer, D. M., S. C. Pratt, H. S. Seidel, and L. Kruglyak. 2006. Population genomic analysis of outcrossing and recombination in yeast. *Nat. Genet.* 38:1077–1081.
- Schurko, A. M., M. Neiman, and J. M. Logsdon. 2009. Signs of sex: what we know and how we know it. *Trends Ecol. Evol.* 24:208–217.
- Schweizer, R. M., J. Robinson, R. Harrigan, P. Silva, M. Galverni, M. Musiani, R. E. Green, J. Novembre, and R. K. Wayne. 2016. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Mol. Ecol.* 25:357–379.
- Scott, M. F., and M. Rescan. 2017. Evolution of haploid–diploid life cycles when haploid and diploid fitnesses are not equal. *Evolution (N. Y.)* 71:215–226.
- Serero, A., C. Jubin, S. Loeillet, P. Legoix-Né, and A. G. Nicolas. 2014. Mutational landscape of yeast mutator strains. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.1314423111.

- Seung, D., A. Echevarría-Poza, B. Steuernagel, and A. M. Smith. 2020. Natural Polymorphisms in Arabidopsis Result in Wide Variation or Loss of the Amylose Component of Starch. *Plant Physiol.* 182:870–881.
- Sharma, U., S. Gupta, S. Venkatesh, A. Rai, A. C. Dhariwal, and M. Husain. 2018. Comparative genetic variability in HIV-1 subtype C p24 Gene in early age groups of infants. *Virus Genes* 54:647–661.
- Sharp, N. P. N. P., L. Sandell, C. G. C. G. James, and S. P. S. P. Otto. 2018. The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc. Natl. Acad. Sci. U. S. A.* 115:E5046–E5055.
- Shoemaker, D. D., D. A. Lashkari, D. Morris, M. Mittmann, and R. W. Davis. 1996. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat. Genet.* 14.
- Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. D. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, doi: 10.1101/gr.3715005.
- Slonimski, P. P., and J. Lazowska. 1977. Transposable Segments of Mitochondrial DNA: A Unitary Hypothesis for the Mechanism of Mutation, Recombination, Sequence Reiteration and Suppressiveness of Yeast “Petite Colony” Mutants. Pp. 39–52 in W. Bandlow, R. J. Schweyen, K. Wolf, and F. Kaudewitz, eds. *Mitochondria 1977: genetics and biogenesis of mitochondria*. Walter de Gruyter & Co., Berlin, Germany., Berlin, Boston.
- Slugina, M. A., A. V. Shchennikova, O. N. Pishnaya, and E. Z. Kochieva. 2018. Assessment of the fruit-ripening-related *FUL2* gene diversity in morphophysiologicaly contrasted cultivated and wild tomato species. *Mol. Breed.* 38:82.
- Smith, C., and D. Greig. 2010. The Cost of Sexual Signaling in Yeast. *Evolution (N. Y.)*. 64:3114–3122.
- Smith, C., A. Pomiankowski, and D. Greig. 2013. Size and competitive mating success in the yeast *Saccharomyces cerevisiae*. *Behav. Ecol.* 25:320–327.
- Sprague, G. F., and I. Herskowitz. 1981. Control of yeast cell type by the mating type locus: I. Identification and control of expression of the a-specific gene *BAR1*. *J. Mol. Biol.* 153:305–321.
- Stage, M., A. Wichmann, M. Jørgensen, N. I. Vera-Jimenéz, M. Wielje, D. S. Nielsen, A. Sandelin, Y. Chen, and A. Baker. 2020. *Lactobacillus rhamnosus* GG Genomic and Phenotypic Stability in an Industrial Production Process. *Appl. Environ. Microbiol.* 86.
- Steden, M., R. Betz, and W. Duntze. 1989. Isolation and characterization of *Saccharomyces cerevisiae* mutants supersensitive to G1 arrest by the mating hormone a-factor. *Mol. Gen. Genet.* 219:439–444.
- Stephan, W. 2019. Selective sweeps. *Genetics*, doi: 10.1534/genetics.118.301319.
- Sun, J., X. Duan, A. A. Hoffmann, Y. Liu, M. R. Garvin, L. Chen, G. Hu, J. Zhou, H. Huang, X. Xue, and X. Hong. 2019. Mitochondrial variation in small brown planthoppers linked to multiple traits and probably reflecting a complex evolutionary trajectory. *Mol. Ecol.* 28:3306–3323.
- Sung, W., M. S. Ackerman, S. F. Miller, T. G. Doak, and M. Lynch. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.1216223109.
- Szafraniec, K., D. M. Wloch, P. Sliwa, R. H. Borts, and R. Korona. 2003. Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast *Saccharomyces cerevisiae*. *Genet. Res.* 82:19–31.
- Takahashi, G., S. Hasegawa, Y. Fukutomi, C. Harada, M. Furugori, Y. Seki, Y. Kikkawa, and K. Wada. 2017. A novel missense mutation of *Mip* causes semi-dominant cataracts in the *Nat* mouse. *Exp. Anim.* 66:271–282.
- Tancos, M. A., A. J. Sechler, E. W. Davis, J. H. Chang, B. K. Schroeder, T. D. Murray, and E. E. Rogers. 2020. The Identification and Conservation of Tunicaminyluracil-Related Biosynthetic Gene Clusters in Several *Rathayibacter* Species Collected From Australia, Africa, Eurasia, and North America. *Front. Microbiol.* 10.
- Taylor, D. R., C. Zeyl, and E. Cooke. 2002. Conflicting levels of selection in the accumulation of mitochondrial

- defects in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 99:3690–3694.
- Thomasson, K. M., A. Franks, H. Teotónio, and S. R. Proulx. 2020. Testing the adaptive value of sporulation in budding yeast using experimental evolution. *bioRxiv*, doi: 10.1101/2020.02.23.959684.
- Thornber, C. S. 2006. Functional properties of the isomorphic biphasic algal life cycle. P. *in* *Integrative and Comparative Biology*.
- Unckless, R. L., and H. A. Orr. 2020. The population genetics of evolutionary rescue in diploids: X chromosomal versus autosomal rescue. *Am. Nat.* 195:561–568.
- Van Bel, M., T. Diels, E. Vancaester, L. Kreft, A. Botzki, Y. Van De Peer, F. Coppens, and K. Vandepoele. 2018. PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.*, doi: 10.1093/nar/gkx1002.
- Van Uden, N. 1985. Temperature Profiles of Yeasts. *Adv. Microb. Physiol.* 25:195–251.
- Vanhoenacker, E., L. Sandell, and D. Roze. 2018. Stabilizing selection, mutational bias, and the evolution of sex*. *Evolution (N. Y.)*, doi: 10.1111/evo.13547.
- Vázquez-Miranda, H., J. A. Griffin, J. M. Sheppard, J. M. Herman, O. Rojas-Soto, and R. M. Zink. 2017. Morphological and molecular evolution and their consequences for conservation and taxonomy in the Le Conte's thrasher *Toxostoma lecontei*. *J. Avian Biol.* 48:941–954.
- Veilleux, C. C., E. E. Louis, and D. A. Bolnick. 2013. Nocturnal light environments influence color vision and signatures of selection on the OPN1SW opsin gene in nocturnal lemurs. *Mol. Biol. Evol.* 30:1420–1437.
- Wagstaff, J. E., S. Klapholz, and R. E. Esposito. 1982. Meiosis in haploid yeast. *Proc. Natl. Acad. Sci.* 79:2986–2990.
- Wang, K., S. Tian, J. Galindo-González, L. M. Dávalos, Y. Zhang, and H. Zhao. 2020. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol. Ecol.* 29:4366–4381.
- Weiller, G. F., H. Bruckner, S. H. Kim, E. Pratje, and R. J. Schweyen. 1991. A GC cluster repeat is a hotspot for mit-macro-deletions in yeast mitochondrial DNA. *MGG Mol. Gen. Genet.* 226:233–240.
- Weissman, D. B., and N. H. Barton. 2012. Limits to the rate of adaptive substitution in sexual populations. *PLoS Genet.*, doi: 10.1371/journal.pgen.1002740.
- Whitlock, M. C. 2002. Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160:1191–1202.
- Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Véronneau, M. Voet, G. Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston, and R. W. Davis. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science (80-.)*, doi: 10.1126/science.285.5429.901.
- Wright, S. 1934. Physiological and Evolutionary Theories of Dominance. *Am. Nat.* 68:24–53.
- Yakubu, A., A. Salako, D. De, and I. Imumorin. 2017. Application of computational algorithms to assess the functionality of non-synonymous substitutions in MHC DRB gene of Nigerian goats. *Genetika* 49:63–76.
- Yoshida, K., A. Ishikawa, A. Toyoda, S. Shigenobu, A. Fujiyama, and J. Kitano. 2019. Functional divergence of a heterochromatin-binding protein during stickleback speciation. *Mol. Ecol.* 28:1563–1578.
- Yoshida, K., T. Makino, and J. Kitano. 2017. Accumulation of Deleterious Mutations on the Neo-Y Chromosome of Japan Sea Stickleback (*Gasterosteus nipponicus*). *J. Hered.* 108:63–68.
- Yoshida, K., R. Miyagi, S. Mori, A. Takahashi, T. Makino, A. Toyoda, A. Fujiyama, and J. Kitano. 2016. Whole-genome sequencing reveals small genomic regions of introgression in an introduced crater lake population of threespine stickleback. *Ecol. Evol.* 6:2190–2204.

- Yoshida, K., M. Ravinet, T. Makino, A. Toyoda, T. Kokita, S. Mori, and J. Kitano. 2020. Accumulation of Deleterious Mutations in Landlocked Threespine Stickleback Populations. *Genome Biol. Evol.* 12:479–492.
- Yoshitomi, H., Y. Ashizuka, S. Ichihara, T. Nakamura, A. Nakamura, T. Kobayashi, and J. Kajiwara. 2018. Molecular epidemiology of coxsackievirus A6 derived from hand, foot, and mouth disease in Fukuoka between 2013 and 2017. *J. Med. Virol.* 90:1712–1719.
- Zepeda Mendoza, M. L., Z. Xiong, M. Escalera-Zamudio, A. K. Runge, J. Thézé, D. Streicker, H. K. Frank, E. Loza-Rubio, S. Liu, O. A. Ryder, J. A. Samaniego Castruita, A. Katzourakis, G. Pacheco, B. Taboada, U. Löber, O. G. Pybus, Y. Li, E. Rojas-Anaya, K. Bohmann, A. Carmona Baez, C. F. Arias, S. Liu, A. D. Greenwood, M. F. Bertelsen, N. E. White, M. Bunce, G. Zhang, T. Sicheritz-Pontén, and M. P. T. Gilbert. 2018. Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat. Ecol. Evol.* 2:659–668.
- Zeyl, C., B. Andreson, and E. Weninck. 2005a. Nuclear-mitochondrial epistasis for fitness in *Saccharomyces cerevisiae*. *Evolution (N. Y.)* 59:910–914.
- Zeyl, C., and G. Bell. 1997. The advantage of sex in evolving yeast populations. *Nature* 388:465–468.
- Zeyl, C., C. Curtin, K. Karnap, and E. Beauchamp. 2005b. Antagonism between Sexual and Natural Selection in Experimental Populations of *Saccharomyces cerevisiae*. *Evolution (N. Y.)* 59:2109–2115.
- Zhang, M., X. Su, E. Mileykovskaya, A. A. Amoscato, and W. Dowhan. 2003. Cardiolipin is not required to maintain mitochondrial DNA stability or cell viability for *Saccharomyces cerevisiae* grown at elevated temperatures. *J. Biol. Chem.*, doi: 10.1074/jbc.M306729200.
- Zhang, M., L. Zhou, R. Bawa, H. Suren, and J. A. Holliday. 2016. Recombination Rate Variation, Hitchhiking, and Demographic History Shape Deleterious Load in Poplar. *Mol. Biol. Evol.* 33:2899–2910.
- Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.1323011111.
- Zörgö, E., K. Chwialkowska, A. B. Gjuvsland, E. Garré, P. Sunnerhagen, G. Liti, A. Blomberg, S. W. Omholt, and J. Warringer. 2013. Ancient Evolutionary Trade-Offs between Yeast Ploidy States. *PLoS Genet.* 9.

Appendices

Appendix A

A.1 Selective markers

Because the *STE5* promoter is active only in haploids, only haploids can grow on uracil drop-out medium. Haploids can also be selected against by the use of 5-fluoroorotic acid (5-FOA), which is converted into a toxic byproduct by the uracil biosynthesis pathway. This means only diploid yeast are able to grow on medium supplemented with 5-FOA. Because the *STE2* promoter is active only in *MATa* haploids, growth on histidine drop-out medium selects for *MATa* haploids. Similarly, the *STE3* promoter is *MATα*-specific, giving these cells the ability to grow on leucine drop-out medium. The antibiotic markers KanMX and HphMX have been inserted close to the *MAT* locus. These markers select for *MATa* through resistance to geneticin (g418) and for *MATα* through resistance to hygromycin, respectively. When the two antibiotics are combined in the medium, only *MATa/MATα* yeast can grow. Spot assays of the haploid lines used in this study, as well as the diploid made from mating the two, on the three different selective media used in this study is shown in Figure A.1.

A.2 Dilutions for mating

For all dilutions, we use dilution factor, $DF = V_f / V_i$, where V_i is the volume of culture and V_f is the total volume of medium with added culture.

In the equal mating ratio treatment, 100 μ L of saturated culture of each mating type was added to the well ($V_i = 100$), after which 800 μ L of medium was added and removed. The 100 μ L of the opposite mating type culture contributes to the total volume and gives $V_f = 1000$ (DF 10). Then, 800 μ L of medium was added ($V_i = 200$, $V_f = 1000$, DF 5). This means, for each mating

type, there was a 50-fold dilution (10×5), while for cell concentration, there was a 25-fold dilution (5×5).

In contrast, in the skewed mating ratio we use 1 mL of culture of the one mating type and 100 μL for the other, then removed 900 μL (a 11-fold dilution for the rare mating type, $V_i = 100$, $V_f = 1100$). We then added 800 μL of medium ($V_i = 200$, $V_f = 1000$, DF 5). This resulted in a 55-fold dilution of the rare mating type and a 5.5-fold dilution for the common mating type (roughly as in the equal mating ratio treatment). There was a 5-fold dilution for overall cell concentration.

A.3 Flow cytometry

Strains were inoculated from 10 μL of freezer stock placed into 1 mL of liquid YPAD medium within 96 well plates. They were grown shaking overnight at 30 °C. 20 μL of the overnight cultures was transferred to 1.5 mL microcentrifuge tubes and washed in 1 mL of sterile water before cells were killed using 1 mL of cold 70% ethanol. The tubes were then kept either at room temperature for at least 1h, or overnight in the refrigerator. After the ethanol treatment, the cells were pelleted at 2500 rpm for 2 minutes, then washed in 1 mL of 50 mM sodium citrate, before being treated with RNAase (25 μL of 10 mg mL^{-1} RNAase added to 1 mL of sodium citrate solution per line). The lines were kept in the RNAase mixture overnight in a 37 °C incubator. The following day, lines were again spun down, supernatant removed, and stained with Sytox Green (30 μL of 50 mg mL^{-1} Sytox Green added to 1 mL of sodium citrate solution). The stained lines were left at room temperature and in darkness overnight.

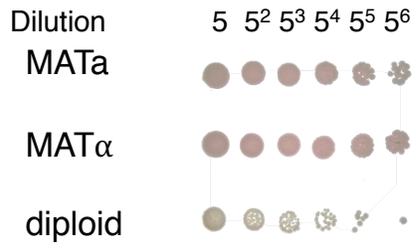
A.4 Phenotyping

Lines were spotted on SE and selective plates: SE+5FoA+g418+hygro for diploid, SE-UH+g418 for *MATa* haploids and SE-UL+hygro for *MAT α* haploids. After two days of growth at 30°C we scored the plates visually. Two lines showed growth on more than one kind of selective medium:

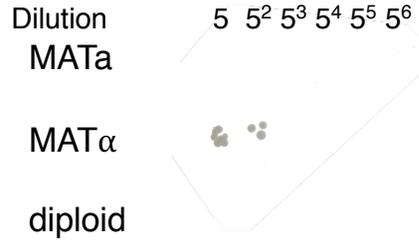
one *MAT α* haploid (line 45) showed weak growth on the diploid selective medium, and one line from the immediate mating treatment (line 177) showed growth both on the diploid selective and on the SE-UL+hygro media. We did not further investigate whether these rare exceptions in growth were due to polymorphism within the cultures, mutational effects, or experimental error.

To test mating type, all lines were spotted on YPAD. This plate was replica plated onto two fresh YPAD plates for mating to the *MAT α* and *MATa* tester. We replica plated a lawn of one of the *MAT* tester onto a velvet and then pressed this lawn onto one of the YPAD plates (containing our lines). We repeated the procedure but for the other mating tester and YPAD plates. After 6h, we replica plated these mating plates onto agar plates with medium lacking eight amino acids (-Arg -His -Leu -Lys -Met -Trp -Ade -Ura). The drop-out plates were left to grow for two days at 30 °C before scoring. While the plates that resulted from mating to the *MATa* tester were easy to score and grew according to expectations, the plates resulting from mating to the *MAT α* tester showed contradictory results (both in comparison to the results from the growth on selective plates as well as mating to the *MATa* tester), suggesting contamination of the tester stock. Because of uncertainty about the *MAT α* tester stock used, we only report results from the *MATa* tester stock.

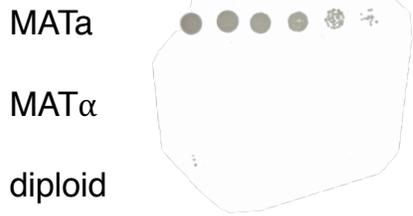
SE media (two days growth)



MAT α -specific (four days growth)



MATa-specific (two days growth)



Diploid-selective (two days growth)

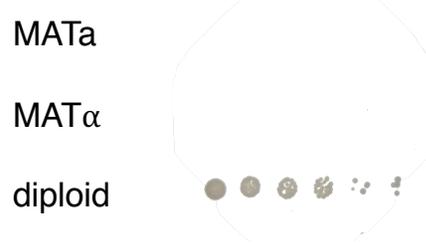


Figure A.1 Spot assay of ancestral lines on selective media

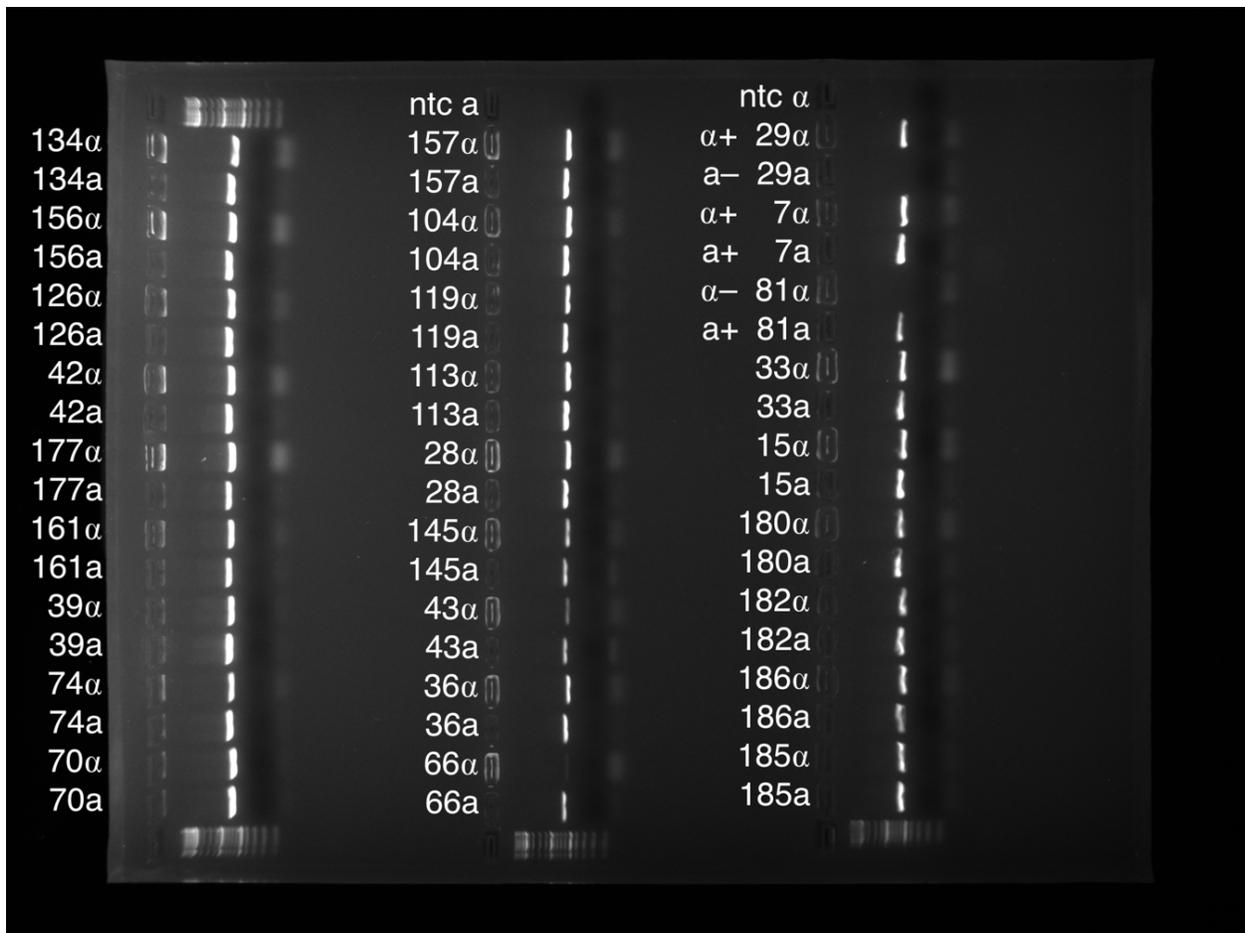


Figure A.2 Results of PCR for *MAT* locus of IM lines. Line ID is followed by the abbreviation a or α depending on which primer was used (see Table A.2). ntc stands for no template control. 43 α and 66 α have faint but existing bands. Line 29 acts as a positive *MAT* α and negative *MAT*a control. Line 7 acts as a positive control for both mating type alleles, and line 81 is our negative *MAT* α and positive *MAT*a control.

Table A.1 Experimental design. The colours in the table designate the type of medium used: permissive SE (green), diploid-specific (pink), presporulation (light blue), sporulation (purple), mating type-specific (yellow). All transfers were done by placing 10 μ L of culture in 1mL of fresh medium.

Day	Evolutionary treatment	
	Delayed mating	Immediate mating
1	Growth (SE)	
2 am ^a	Mating	Growth (SE)
pm	Diploid selection	
3 am	Diploid selection	
pm ^a	Presporulation	
4	Sporulation	
5	Sporulation	
6 am	Ascus digestion	
pm	Germination (<i>MAT</i> -specific)	Germination (SE)
7	<i>MAT</i> -specific selection	Diploid selection

^a Before mating and before transfer to presporulation medium, 200 μ L of the lines were sampled for storage, until the end of the experiment when frozen (adding glycerol to a final concentration of 15 %).

Table A.2 Primers used in genotyping *MAT* locus

Primer	Purpose	Location	Sequence
OLPr005	Forward <i>MAT</i> α	1884 bp into <i>MAT</i> locus	AAGTTGCAAAGAAATGTGGC
OLPr006	Forward <i>MAT</i> α	2150 bp into the <i>MAT</i> locus	AAAATGCAGCACGGAATATG
OLPr007	Reverse (common)	217 bp upstream of the <i>MAT</i> locus	AACAAATTGTGAAGCCGAAG

Appendix B

Table B.1 Number of petite and grande lines of each ploidy and *RDH54* type in DS1

Ploidy	Control or MA	Not petite		Petite	
		<i>RDH54</i>	<i>rdh54Δ</i>	<i>RDH54</i>	<i>rdh54Δ</i>
Haploid	Control	17	16	0	0
	MA	45	46	5	2
Diploid	Control	6	21	28	11
	MA	17	47	82	48

Table B.2 Number of petite and grande lines of each ploidy and *RDH54* type in DS2

Ploidy	Control or MA	Not petite		Petite	
		<i>RDH54</i>	<i>rdh54Δ</i>	<i>RDH54</i>	<i>rdh54Δ</i>
Haploid	Control	11	8	6	8
	MA	21	24	28	24
Diploid	Control	7	22	27	10
	MA	12	57	86	38