GENOME-WIDE INVESTIGATIONS INTO THE CONSEQUENCES OF GENE AND CHROMOSOME COPY NUMBER CHANGES

Abigail N. Keller

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Maitreya J. Dunham, Chair

Stanley Fields

Matt R. Kaeberlein

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Genome-wide investigations into the consequences of gene and chromosome copy number changes

Abigail N. Keller

Chair of the Supervisory Committee:

Maitreya J. Dunham, Professor

Departments of Genome Sciences

Aneuploidy can have contradictory effects on organisms and cells, from developmental disorders and cancer in humans to drug resistance and adaptive advantage in yeast. Despite a long history of involvement in human health and innumerable cases in other model organisms, how aneuploidy confers seemingly paradoxical effects on cellular fitness remains debated. In this work, we examined the contributions individual genes may have on the fitness of aneuploid strains in environmental conditions normally detrimental to growth. We used a collection of ~1,800 *Saccharomyces cerevisiae* strains with large amplifications tiled across the genome to look at how aneuploidy impacts fitness in high temperature, prolonged stationary phase, and treatment with the Hsp90 inhibitor radicicol, as well as a normal temperature and DMSO only control. This screen revealed 91 candidate regions that may contain single genes that

disproportionately impact fitness when amplified. Our candidate regions were identified using a new method to fit a piecewise constant model to fitness data across each chromosome arm. To validate potentially impactful genes, we developed a method to systematically test the contribution of each candidate gene in a pooled fashion. Finally we investigated the causes and consequences of the environmental stress response (ESR), a gene-expression signature exhibited by aneuploid cells. By controlling the growth rate of complex aneuploid strains, we show that the ESR is dependent on the proliferation rate of cells and results in decreased cell density of aneuploids via ribosome loss.

Table of Contents

List of Figures

List of Tables

Acknowledgments

The work in this thesis would not be possible without the incredible mentors and science teachers I have had throughout my academic career. My independent study in high school with Jason Anderson first sparked my interest in evolution and inspired me to major in biochemistry at the University of Wisconsin – Madison. I would particularly like to thank my first lab mentors Yevgenya Grinblat and Jessica TeSlaa, who taught me the ins and outs of benchwork and being part of a lab and gave me my first experience in working on a published manuscript. At the University of Washington I was fortunate enough to find my way to the lab of Maitreya Dunham, who has been incredibly supportive with an endless excitement for genetics that has always been inspiring. Thank you to Maitreya for both her mentorship and her cultivation of a close-knit and fun lab environment. Thank you as well to former and current members of the Dunham Lab, particularly to Anna Sunshine, whose work was the basis for much of this work, and Celia Payen, who first mentored me in the lab and showed me the ropes. Thank you to my committee members, Bonny Brewer, Sue Biggins, Stan Fields, and Matt Kaeberlein for thoughtful discussion, advice, and support throughout graduate school. I have made an amazing group of friends in my cohort, particularly Brittany Whitley, Sophie Archambeault, and Becca Martin, who were always there to celebrate, commiserate, and support me and each other. Lastly, I would like to thank my family for their endless support, particularly my parents, Betty and Joel Keller, my brother Matt Keller, my grandmother Betty Halverson, and my aunt Robbie Keller for always reminding me that I can accomplish whatever I put my mind to.

Chapter 1. Introduction

Faithful replication and segregation of the cell's genome is one of the most important processes of life. Errors in this process can have catastrophic effects on an organism's ability to survive and propagate. Aneuploidy, or an abnormal number of chromosomes, is just one way errors in mitosis and meiosis can manifest in a cell. This type of genomic abnormality can result in growth delays, disease, and even death. However, aneuploidy can also serve as a route to adaptation in populations experiencing environmental stress. And while we are becoming increasingly aware of the widespread effects aneuploidy has on disease, aging, and even normal tissue types, the mechanism by which aneuploidy influences cellular fitness is still under much debate.

1.1 Definition and causes of aneuploidy

Aneuploidy is defined as an aberrant number of whole chromosomes, specifically an integer number differing from a multiple of the haploid complement. This can encompass an increase or decrease in chromosome number, though this difference usually affects only one or very few chromosomes, as larger changes are often lethal. For example, a diploid cell with an extra copy of one chromosome is trisomic (a base haploid would be disomic), while a diploid cell with one missing chromosome is monosomic. This differs from the change in copy number of a full haploid complement of chromosomes, known as polyploidy, in that it results in an unbalanced number of chromosomes. While aneuploidy strictly refers to the gain or loss of whole chromosomes, or in some cases full chromosome arms, in this thesis I will be using the term 'aneuploidy' to encompass all large-scale amplifications or deletions, both chromosomal and segmental, for simplicity.

Whole chromosome aneuploidy is usually the result of errors in mitosis or meiosis, the mechanism by which cells replicate and divide. Chromosome segregation is tightly regulated by the spindle assembly checkpoint, ensuring that paired chromosomes remain together until they are properly aligned on the spindle (reviewed in Lara-Gonzalez et al. 2012). And while normally rare— potentially as low as 1 in 10,000 meiotic divisions in *Saccharomyces cerevisiae* (Sears et al. 1992)— missegregation events do happen. Chromosome nondisjunction, the failure of homologous chromosomes to separate appropriately in meiosis I or sister chromatids in meiosis II and mitosis, will result in an abnormal number of chromosomes to be inherited by daughter cells. In meiosis, this can lead to all daughter cells, or a whole organism, bearing a chromosome abnormality. Defects in components of the spindle assembly checkpoint are not the only errors that can disrupt this process. Replication stress, the slowing or stalling of replication forks leading to a delay in S-phase, has been shown to be a major cause of chromosomal instability (CIN) leading to aneuploidy (reviewed in Wilhelm et al. 2020).

Segmental aneuploidies can also arise through a number of different mechanisms. Diverse intra- and interchromosomal segmental duplications have been found to arise in yeast at an estimated rate of 10^{-9} per cell per division (Koszul et al. 2004; Koszul et al. 2006). Defects during chromosome segregation result in lagging chromosomes, but studies show that most lagging chromosomes are not missegrated, instead ending up in micronuclei in the daughter cell (Thompson and Compton 2011). These micronuclei can accumulate massive DNA damage and cause fragmentation and rearrangements that are sometimes reincorporated into the nucleus (Crasta et al. 2012). Damaged chromosomes from micronuclei can also induce chromothripsis, cycles of copy number changes and structural rearrangements of one or many chromosomes (Zhang et al. 2015). Structural changes also arise in the genome as the result of unbalanced

translocations. This is relatively common in well studied yeast species and repetitive elements such as retrotransposons (Ty elements), long terminal repeats (LTRs), and tRNAs are all capable of serving as breakpoints for translocations (Lemoine et al. 2005, Chan and Kolodner 2011, Gresham et al. 2008). When cells with unbalanced translocations replicate, daughter cells will inherit a gain or loss of the translocated chromosome segment.

1.2 Aneuploidy and human health

Aneuploidy is, by definition, an unbalanced number of chromosomes. Because expression levels largely correlate with gene copy number in the absence of dosage compensation mechanisms, this results in a large number of gene products expressed at an abnormal level. In single-celled organisms these imbalances can sometimes be tolerated, but in humans and other mammals they are often highly deleterious. Aneuploidy is the most common cause of spontaneous abortion and congenital defects in humans, accounting for 35% of pregnancy losses (reviewed in Nagaoka et al. 2012). Furthermore, only three trisomies are capable of surviving beyond the first weeks of infancy— chromosome 13 (Patau Syndrome), chromosome 18 (Edwards Syndrome), and chromosome 21 (Down Syndrome). Beyond whole chromosome abnormalities on an organismal level, aneuploidy in specific tissues has been linked with cancer and other diseases and disorders in humans.

1.2.1 Aneuploidy and cancer

Chromosome copy number and structural abnormalities are so prevalent among cancers that aneuploidy is often referred to as a characteristic of cancer cells. According to the Mitelman Database for Chromosome Aberrations in Cancer, 90% of solid tumors and 50% of hematological neoplasms are aneuploid (Simonetti et al. 2019, Mitelman Database 2017). Links between aneuploidy and cancer have been hypothesized for almost a century, with David

Hansemann writing in 1890 of his observations of asymmetrical nuclear divisions in epithelial cancer cells and Theodor Boveri postulating in 1902 that these abnormal divisions produced chromosome combinations that may induce tumorigenesis (Boveri 2008; Hardy and Zacharias 2013).

Despite this long history, the debate over whether aneuploidy itself is enough to cause cancer or merely arises as a consequence of unchecked cell growth remains contested. The overwhelming presence of aneuploidy among various cancer types suggests that these genomic alterations provide some benefit. However, experimentally introducing an extra chromosome can actually have a tumor suppressing effect as shown in mouse xenograph models (Sheltzer et al. 2017). Cells had to undergo further karyotype rearrangements to improve cellular fitness, indicating single chromosome changes are not sufficient to promote tumorigenesis. Additionally, the characteristic phenotypic responses seen in cell models of aneuploidy, namely metabolic and proteotoxic stress signatures, are also present in primary tumor cells (Dai and Sampson 2016, Sheltzer 2013). And tumor suppressing effects of aneuploidy can be observed in the rates of cancer seen in humans with aneuploidy syndromes. While individuals with Down Syndrome have elevated risk of acute myelocytic leukemia (AML), an overview of 17,897 individuals over 15 years found all other malignant neoplasms as cause of death less than one-tenth as often as expected (Yang et al. 2002).

In contrast, there is evidence that aneuploidy may precede cancer transformation rather than merely arising after the fact. Aneuploidy has been observed to be predictive of poor outcomes in several cancer types. Even when controlled for occupational exposure to carcinogens and smoking habits, chromosome aberrations in peripheral blood lymphocytes (PBLs) were found to be predictive of high risk for developing leukemia (Bonassi et al. 2000).

In prostate cancer, aneuploidy was found to be predictive of lethal outcomes even when controlled for Gleason score, the most common histological grading system for prostate cancer (Stopsack et al. 2019). This trend applies to a wide range of cancer types. A study of 18 gene expression data sets, markers of chromosome instability were significant predictors of clinical outcome in two-thirds, representing glioma, lung cancer, lymphoma, medulloblastoma, and mesothelioma (Carter et al. 2006).

The answer to why aneuploidy is both detrimental to cellular fitness and so overwhelmingly present in highly proliferative cancer cells may lie in the specific karyotype of the cell and the environment in which the tumor arises. Cancers can be thought of as an evolving population of cells where the fastest growing individuals grow and spread to take over the tumor. And at least some aspect of the successful cell's karyotypes must provide a growth benefit compared to nearby cells. Even specific gene mutations, like *BRCA1/BRCA2* mutations linked to breast cancer, can have different consequences in different tissues. This extends copy number variations across cancer subtypes. Different cancers have vastly different karyotypes that recur compared to cancers originating in other parts of the body (Kim et al. 2013). An increased copy number of a chromosome in one type of tumor may almost never be seen in another. A meta-analysis of copy number profiles of over 8000 cancer genomes found that hierarchical clustering of frequencies of chromosome arm gains and losses was able to cluster cancer types by developmental lineages, highlighting the tissue specific nature of the cancer promoting effects of aneuploidy (Kim et al. 2013). This is highly suggestive that the specific genes present on particular chromosomes are critical to whether it will promote rapid growth in a specific environment, in this case a specific organ or tissue type.

1.2.2 Aneuploidy in human disorders

We are becoming increasingly aware of the role that previously unidentified structural and copy number variants play in human health and disease. Due to the highly deleterious nature of chromosome level copy number changes in humans, most large scale changes to the genome present with developmental and mental disability (reviewed in Thiesen and Shaffer 2010). Smaller scale deletions and duplications on the order of kilobases to megabases, however, previously went unnoticed prior to more widespread genome sequencing efforts. As human genome sequencing becomes more widely available, we are able to identify these types of mutations, though the complex nature of large genomic alterations makes understanding their functional impact difficult.

CNVs can result in phenotypic changes via several different mechanistic routes. The most obvious path to phenotypic effects is the change in copy number of one or many dosage-sensitive genes. In some cases, particularly when the region affected is small, it is relatively easy to trace phenotype back to a specific gene. One example of this is illustrated in the well-studied case of Charcot-Marie-Tooth Type 1a (CMT1a). The roughly 1.5 Mb tandem duplication of chromosome 17p12 impacts several genes, though duplication of the gene *PMP22* is widely accepted to cause the symptoms associated with CMT1a (Chance et al. 1994, Inoue et al. 2001). Dosage sensitivity in *PMP22* also plays a role in the related deletion of the same region found in the disease hereditary neuropathy with liability to pressure palsy (HNPP) (Lupski et al. 1992).

In other cases, the presence of a large number of genes on the deleted or duplicated segment complicates causative links. For example, Williams Syndrome is a neurodevelopmental disorder characterized by deletion of between 26 and 28 genes on chromosome 7. Changes in copy number at these loci together result in the distinctive physical, developmental, and

behavioral features of the disorder, though links between specific genes and phenotypes, such as haploinsufficiency of *ELN* and characteristic heart defects, are still largely unclear for many of the affected genes (Chailangkarn et al. 2018). Other consequences of deletion syndromes can include unmasking of a disease associated allele. Sotos syndrome, a disorder largely associated with deletion or disruption of *NSD1*, often presents in the context of a larger microdeletion (Tatton-Brown et al. 1993). An investigation of several patients with low coagulation factor twelve (FXII) activity found that this symptom was the result of a deletion including the neighboring *FXII* gene and the phenotype was predominantly dependent on the functional activity of the remaining allele (Kurotaki et al. 2005).

Copy number variants are increasingly being found as potential contributors to more complex neurological disorders. Autism spectrum disorders are neurodevelopmental disorders presenting with social, behavioral, and communication symptoms. As a behaviorally defined group of disorders not linked to any one specific gene, ASDs have been associated with a variety of genetic factors including structural and copy number variants (Zarrei et al. 2019, Sebat et al. 2007, Marshall et al. 2008, Mefford and Eichler 2009). The number of regions potentially associated with increased risk of ASD and their incomplete penetrance highlights not only the possibility that many genes contribute to observed phenotypes, but also the prospect that complex gene-environment interactions may play a role (Koufaris and Sismani 2015). A similar situation exists for genetic determinants of schizophrenia and epilepsy (Ahn et al. 2014, Helbig et al. 2014). In fact, it has been hypothesized that some CNVs with low penetrance for schizophrenia may still be highly pathogenic, but manifest as other neurodevelopmental disorders such as developmental delay or ASDs (Kirov et al. 2014).

Despite clear associations between particular CNVs and heightened risk of neurodevelopmental and psychiatric disorders, narrowing down phenotypes to particular genes has again been difficult. CNVs can be completely benign or highly pathogenic depending on genetic content. It has been shown that more pathogenic CNVs are enriched for developmental genes and genes that are more dosage-sensitive, and less pathogenic CNVs contain genes with naturally more variable copy numbers (Rice and McLysaght 2017). Work has also been done to determine whether multiple hits are needed for disease phenotypes to arise (Bassett et al. 2017), highlighting the importance of how large amplifications and deletions interact with other mutations in the genome.

1.3 Aneuploidy as an adaptive mechanism

Aneuploidy is naturally occurring in certain species, particularly in single-celled organisms, and can be tolerated or even beneficial in some cases. Extensive genome plasticity in *Leishmania* parasites has been observed, with high levels of aneuploidy observed across strains and species (Rogers et al. 2011, Bastien et al. 1992). The high degree of genomic heterogeneity prevalent in *Leishmania* populations is thought to be a potential mechanism for adaptation to changing environments and emergence of drug resistance (Sterkers et al. 2012, Ubeda et al. 2008, Leprohon et al. 2009). Another human pathogen, the yeast *Candida albicans*, also employs genomic rearrangements as an adaptive strategy for drug resistance (Perepnikhatka et al. 1999, Selmecki et al. 2006, Selmecki et al. 2009). In one study, nearly 50% of *C. albicans* strains found to be resistant to fluconazole had at least one incident of aneuploidy (Selmecki et al. 2006). Further analysis of the chromosome 5L amplification found in a fifth of these cases revealed resistance was due to increased copy number of *ERG11* and *TAC1*, the target of

fluconazole and a transcription factor responsible for upregulating relevant efflux pumps, respectively (Selmecki et al. 2008).

The budding yeast *Saccharomyces cerevisiae* is similarly prone to aneuploidy during the course of adaptation. Laboratory evolution experiments in yeast repeatedly result in aneuploid cells that can readily outcompete ancestral clones (Gresham et al. 2008, Dunham et al. 2002). Specific aneuploid events arise to high frequency in populations grown in nutrient-limited continuous culture. In sulfate-limitation, evolved populations almost always show an elevated copy number of the right arm of chromosome 2 (Gresham et al. 2008). Extensive studies of this trend have revealed that the benefit of this amplification is almost exclusively due to the increased copy number of *SUL1*, a high affinity sulfate transporter (Payen et al. 2014, Sunshine et al. 2015). A similar, though less penetrant, trend is observed in glucose-limitation with amplification of the glucose transporters *HXT6/7* (Dunham et al. 2002) and in nitrogen-limitation with amplification of *GAP1* via extrachromosomal DNA circles (Gresham et al. 2010). These cases demonstrate that aneuploidy often occurs as a mechanism for organisms to adapt to a challenging environment.

1.4 Construction of the Telomeric Amplicon 'Tamp' pool

Much of the work presented and discussed in this thesis involves the use of the telomeric amplicon or 'Tamp' collection of yeast. This collection was developed as part of the thesis work of Anna Sunshine, a former graduate student in the Dunham Lab. The original study using this collection was published in PLoS Biology in 2015 (Sunshine et al. 2015). Many specific attributes of the collection will become relevant in later chapters, so I have provided an overview of the important aspects of construction of the pool here. Previous work had found that of 36 aneuploid events observed in evolved clones from nutrient-limited experiments, 17 were

telomeric amplicons (Gresham et al. 2008, Sunshine et al. 2015). Furthermore, a survey of 4.934 cancers from The Cancer Genome Atlas Pan-Cancer data set found that somatic copy number alterations that were bounded on one side by a telomere were both longer and more frequent than internal duplications, highlighting the importance of telomeric events in understanding the evolution of cancer (Zack et al. 2013).

To create a collection of telomeric amplicon strains tiled across the genome, the heterozygous yeast knockout collection was used as the backbone for strain construction. Strains in this collection all bear a common *KanMX* marker at specific barcoded sites in the genome, each replacing a particular gene of interest (Tong et al. 2001). As the Tamp pool was originally developed for studies in nutrient-limited chemostats, a subset of strains with near-neutral fitness in phosphate-, glucose-, and sulfate-limited media were selected for transformation (Payen et al. 2016). When expanding the use of the Tamp collection to other conditions, we did have to take confounding effects of the background strains into consideration; however, the number of additional strains that needed to be censored was ultimately quite small (see Chapter 2, Supplementary Table 2.3).

Once background strains were collected, they were split into two pools based on the orientation and location of their deletion cassette— genes located on the Watson strand of the left arm or the Crick strand of the right arm (wlcr pool), and genes located on the Watson strand of the right arm or the Crick strand of the left arm (wrcl pool). This allowed for a single region of homology to be used for each of the two pools including a segment of the *KanMX* cassette and the distal-most barcode of the deletion insert. This amplicon was cloned into a chromosome fragmentation vector that contained a telomere-seeding sequence, centromere, *URA3* selectable marker, and the region of homology. When yeast are transformed with a linearized chromosome

fragmentation vector, the DNA fragment acts as a chromosome that has undergone a double-stranded break. The region of homology on the native chromosome acts as a template for repair, creating a new chromosome that starts at the site of the deletion cassette in the cell and extends to the telomere on the same arm of the native chromosome.

While a powerful tool for investigating fitness effects of aneuploidy in many different environmental conditions, the Tamp pool has several key limitations. First, the backbone of the pool is the heterozygous deletion collection. This allowed for a collection of ~1,800 unique karyotypes to be constructed in just two pooled transformations; however, more recent studies have highlighted unintended mutations and chromosomal abnormalities present in the knockout collection (Puddu et al. 2019). Second, subsequent whole-genome sequencing of individual strains from the Tamp collection have revealed that not all strains have the karyotype expected if break-induced replication had extended completely to the telomere. Our construction and analysis approaches were designed to limit the impacts of erroneous Tamp strains by including multiple replicate strains per desired karyotype and filtering out Tamps with high variability between replicates. Finally, the method of identifying breakpoints has changed since the original publication of this collection (see Chapter 2, Sunshine et al. 2015). The original approach to calling candidate breakpoints employed the R package "DNAcopy" and was based on the observation of a step-like pattern mimicking DNA copy number data in the fitness plots of Tamps in nutrient-limited chemostat growth. For our analysis in an expanded set of conditions, we wanted to develop a method that incorporated the possibility of a linear slope in our fitness data. An expanded discussion of the development of this method and potential opportunities for improvement are discussed in Chapter 2.

1.5 Specific Aims

The work presented here is centered around my interest in how aneuploidy can confer such paradoxical effects in different environments. Aneuploidy is widely known to result in severe developmental disorders and even death in humans, yet it is rampant in cancer cells. While cancer is undoubtedly detrimental to an organism's survival, the prevalence of aneuploidy in such rapidly proliferating cells must hint at a proliferative advantage to at least some abnormal karyotypes. In single-celled organisms like budding yeast, aneuploidy results in growth delays and reduced viability, yet often arises in populations adapted to stressful environments. Previous work in the Dunham Lab has sought to understand how aneuploidy may provide an adaptive benefit to cells. In my work, I have expanded our focus to conditions normally more poorly tolerated by aneuploid cells in order to explain how aneuploidy confers a fitness cost.

1.5.1 Effects of large amplifications on fitness in detrimental conditions

The effects of aneuploidy on cellular fitness and cellular processes is an active field of research. These studies have revealed several conditions to which aneuploid yeast show an increased sensitivity. To investigate whether individual genes are responsible for these phenotypes, I used a collection of yeast, each bearing a unique large amplification tiled across the genome, developed by a former student in the lab. I competed this collection of strains in conditions that aneuploid cells tolerate poorly— high temperature, extended stationary phase, and treatment with the Hsp90 inhibitor radicicol. From these experiments I gathered fitness data from ~1,500 Tamp strains across three experimental conditions and two control conditions. With these data, I worked with collaborators to develop a method to identify regions that have a disproportionate effect on fitness when amplified by fitting a piecewise constant model to our fitness data across chromosome arms. This analysis resulted in 91 candidate regions that we can

use to further investigate whether single genes can drive fitness consequences in these conditions.

1.5.2 Validation of impactful genes on Tamp fitness

After identifying candidate regions that may contain particularly impactful genes on the fitness effects we observed in various conditions, we have been developing methods to assay the contribution of individual genes at each of the candidate sites. We identified a region of chromosome 10 that displayed fitness 'downsteps', regions that reduced fitness of the amplification when amplified, in both phosphate- and glucose-limited chemostat growth at proximal but non-overlapping sites. I constructed three large Tamp strains in a haploid background, and mated each to a pool of strains bearing deletions of single genes across these two step points, thereby systematically returning each gene to a euploid copy number. Throughout this process, we optimized a method of 'rescue-by-mating' via random spore analysis by which essential gene deletions were newly sporulated from heterozygous diploids and immediately mated to the Tamp strains before they died from the loss of an essential protein. Though our competition of this pool yielded fitness data from most strains across both breakpoints, results from this analysis were largely inconclusive, possibly from underlying deletion collection strain characteristics or a lack of a single gene driving fitness effects.

1.5.3 Translational capacity of aneuploid yeast in phosphate-limited chemostat culture

Aneuploid cells have been shown to exhibit a particular set of physiological responses to the stress that an unbalanced genome places on the cell. These effects involve changes to metabolism, cell wall integrity, and in particular, a transcriptional response impacting the expression of ~900 genes. This response has been termed the environmental stress response (ESR) and is correlated with the degree of aneuploidy of the cell. To uncouple the effects of

growth rate and aneuploidy on the ESR, we grew complex aneuploid strains in phosphate-limited chemostats to equalize growth rate between aneuploid cells and euploid controls. When growth rate is constant, the correlation between ESR and degree of aneuploidy disappears. Additionally, we found that while aneuploid cells usually show a depletion of ribosomes in relation to euploid cells, this trend also was eliminated when growth rate was controlled. This suggests that ribosome loss in aneuploids is a result of the ESR mounted by the cell due to slow proliferation.

Chapter 2. Effects of large amplifications on fitness in detrimental conditions

---

Work described here was performed by Abigail Keller and collaborators. The Tamp collection was constructed by Anna Sunshine and published in PLoS Biology in 2015 (Sunshine et al. 2015). Noah Hanson performed Nextera Sequencing on provided samples. Lucy Gao, formerly in the Witten Lab at University of Washington, wrote original R scripts to predict fitness breakpoints using the 1D Fused Lasso. This chapter is being prepared as a manuscript for future submission.

Whole-chromosome aneuploidy and large segmental copy number variants (CNVs) can have devastating effects in multicellular organisms, from developmental disorders and miscarriage to cancer. Aneuploidy in single-celled organisms such as yeast also results in proliferative defects and reduced viability. Yet, paradoxically, CNVs are routinely observed in laboratory evolution experiments with microbes in stressful growth conditions. The defects associated with aneuploidy are often attributed to the imbalance of many differentially expressed genes on the affected chromosomes, with many genes each contributing incremental effects. An alternate hypothesis is that a small number of individual genes are large effect 'drivers' of these fitness changes when present in an altered copy number. To test these two views, we have engineered a library of strains with ~1,800 synthetic chromosome arm amplifications tiled across the genome such that each amplification extends from a barcoded genomic location to the telomere on the same arm, and neighboring amplifications differ by only a few genes on average. This collection has previously been assayed under nutrient-limited conditions known to select for aneuploidies. In this work, we focus on conditions known to be poorly tolerated by aneuploid yeast— high temperature, treatment with the Hsp90 inhibitor radicicol, and growth in extended

stationary phase. To identify potential genes with a large impact on fitness, we fit a piecewise constant model to fitness data across chromosome arms, filtering breakpoints in this model by magnitude to focus on regions with a large impact on fitness in each condition. While fitness generally decreased as the length of the amplification increased, we were able to identify 91 candidate regions that disproportionately impacted fitness when amplified. Consistent with our previous work with the Tamp collection, nearly all candidate regions were condition specific, with only five regions impacting fitness in multiple conditions.

2.1 Introduction

Aneuploidy, the presence of an abnormal number of chromosomes in a cell, can have contradictory effects on cellular fitness depending on environmental context. In multicellular organisms, aneuploidy can result in serious medical consequences such as sterility, intellectual disability, neuropsychiatric disorders, and spontaneous abortion (Yurov et al. 2018; Rutkowski et al. 2017; Potapova and Gorbsky 2017). Whole chromosome aneuploidies are often lethal in vertebrates, with only humans carrying autosomal trisomies 13 (Patau Syndrome), 18 (Edwards Syndrome), and 21 (Down Syndrome) capable of survival beyond the first weeks of infancy. But while aneuploidy can be devastating on an organismal level, these same chromosomal abnormalities may contribute to the highly proliferative nature of cancer cells. Aneuploidy, along with chromosome instability (CIN), is extensively seen in tumor cells, to the point of being considered a hallmark of cancer, occurring in nearly 90% of solid tumors and 75% of blood cancers (Weaver and Cleveland 2006; Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer, https://mitelmandatabase.isb-cgc.org/). The tumorigenic nature and proliferative success of certain cancers may result from aneuploidies that alter expression of

tumor suppressors or oncogenes, though causal explanations for this link are still limited in number (Lengauer et al. 1998; Williams et al. 2008; Sotillo et al. 2007).

While it is still in debate whether aneuploidy is a cause or consequence of unchecked cell growth in cancer, recent studies have reported links between the presence of aneuploidy and poor prognosis in cancer patients. Aneuploidy and CIN are associated with a higher chance of lethality in prostate (Stopsack et al. 2019) and breast cancers (Oltmann et al. 2017), as well as invasive behavior in drosophila models of epithelial cancers (Benhra et al. 2018). Additionally, colorectal carcinomas tend to have a characteristic chromosome 7 amplification which is associated with enrichment in signaling pathways crucial for malignant transformation (Braun et al. 2019). Despite the correlation between prognosis and ploidy, it remains unclear whether tumor malignancy is dependent on simply the state of aneuploidy or the specific genes that have an altered copy number.

Single-celled organisms also show a complex relationship between aneuploidy and fitness. Cells with large CNVs are known to exhibit a slower growth rate in rich media, yet CNVs regularly arise under strong selection in stress environments. Budding yeast bearing chromosome abnormalities routinely evolve and outcompete euploid ancestors in nutrient-limiting experiments (Dunham et al. 2002; Gresham et al. 2008), and at least one chromosomal rearrangement has been directly linked to resistance to the antifungal drug fluconazole in *Candida albicans* (Selmecki et al. 2006). *Candida* and *Leishmania* species, both well-known human pathogens, display a high degree of genomic plasticity that may influence drug resistance and infectivity (Selmecki et al. 2005; Laffitte et al. 2016). While more cases of CNVs linked to disease, drug resistance, and pathogenicity continue to be discovered, in very few instances do we completely understand the mechanism by which large or complex chromosomal aberrations

confer a fitness benefit or detriment. The effects of specific aneuploidies have in some cases been pinpointed to causative genes, such as fluconazole resistant *C. albicans* driven by the amplification of *ERG11* and *TAC1* on chromosome 5L (Selmecki et al. 2008) and increased lysine usage in a mutualistic microbial community (Hart et al. 2019), but the majority of cases in both single- and multi-celled organisms are complicated by the sheer number of genes present in an abnormal copy number and the interactions they may have with one another. For example, candidate gene approaches have been slow in finding causative loci for many CNV-disease associations, and studies are biased towards candidate genes with already known or hypothesized links to the disease of interest (Usher and McCarroll 2015).

In our prior work, we addressed this problem in a yeast model by engineering a set of synthetic amplifications tiled across the entire *S. cerevisiae* genome (Sunshine et al. 2015). The strains in this telomeric amplicon, or "Tamp", collection all contain a single amplification starting at a barcoded gene knockout from the heterozygous deletion collection and extending to the telomere of the same chromosome arm (Figure 2.1A). The tiled nature of the Tamp collection revealed a step-like pattern when plotting the fitness of each strain by the chromosomal coordinate of the amplification start site, indicating that the fitness effects of amplifications change very little until particular regions of the chromosome are included in the amplification. This result is consistent with the hypothesis that in a growth condition under which specific large amplifications are beneficial, fitness effects can be conferred by a small number of genes that have been amplified. We revealed that adaptation to nutrient-limited environments through chromosomal copy number alterations was, in fact, largely driven by a discrete number of driver genes.

While this result is consistent with previous studies tracing the fitness benefit of amplifications back to specific genes in nutrient-limited environments or in the presence of antifungal drugs (Payen et al. 2014, Gresham et al. 2008, Dunham et al. 2002, Linder et al. 2017, Selmecki et al. 2008), these are conditions in which certain large amplifications are known to provide an adaptive benefit. In contrast, disomic yeast have been shown to have a reduced fitness in high temperature, extended stationary phase, and treatment with Hsp90 inhibiting drugs (Torres et al. 2007). Aneuploid yeast are seen to share a set of common phenotypes such as increased glucose uptake and sensitivity to protein turnover and homeostasis perturbation due to the increased burden of expressing extra genes on large amplifications (Torres et al. 2007, Terhorst et al. 2020). Here we expand our Tamp screen to include conditions that aneuploid yeast tolerate poorly to investigate whether single genes disproportionately drive fitness effects in detrimental conditions, or if fitness deficits are largely due to the excess expression of many amplified genes.

2.2 Results

2.2.1 Leveraging a genome-wide collection of Telomeric Amplicon strains

Amplification of a large chromosomal segment simultaneously changes the copy number and expression level of a large number of genes. In most cases, it is unclear how many of these genes contribute to the fitness effect of the amplification in a particular growth condition, or how this subset might change in a different environment. We have observed large amplifications rising to high frequency in laboratory evolution experiments under nutrient-limited environments, indicating that specific chromosome rearrangements and copy number changes may be beneficial to growth in these conditions (Gresham et al. 2008, Dunham et al. 2002, Kvitek and Sherlock 2011, Payen et al. 2014). Our previous work with the Tamp pool provided

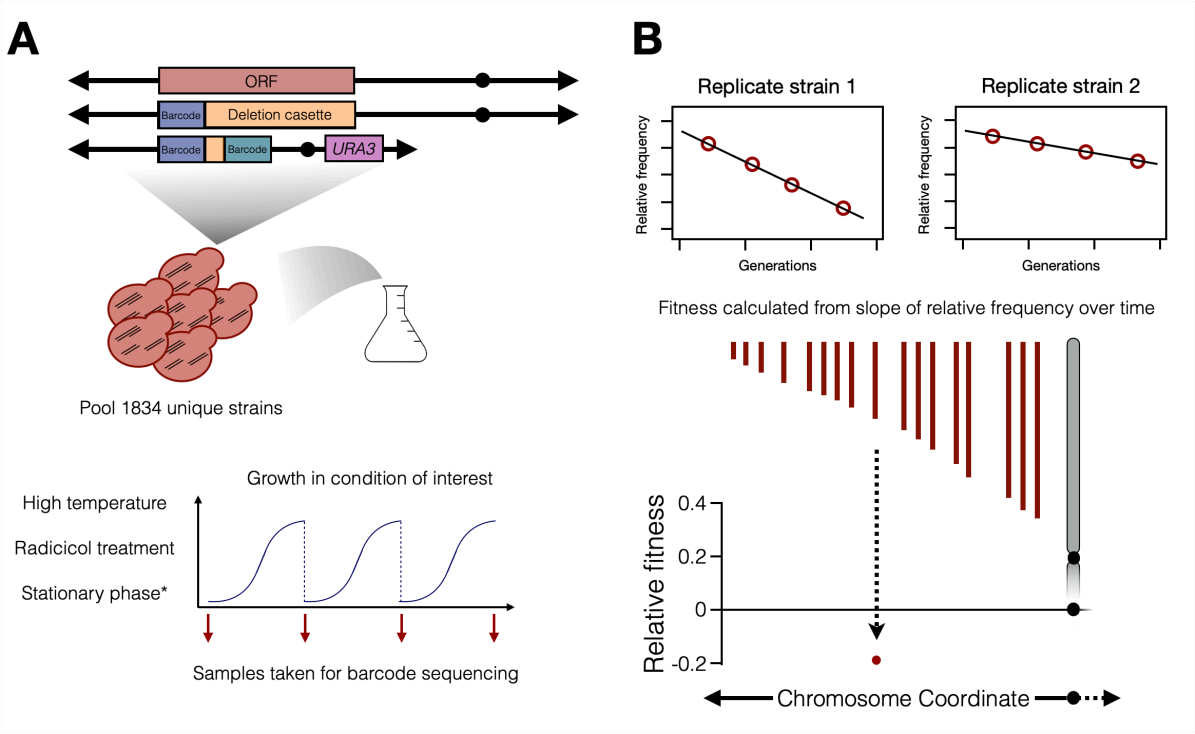


Figure 2.1: Overview of Tamp collection and competitions. (A) Overview of Tamp pool and competition experiments. Tamp strains were constructed by introducing a chromosome fragmentation vector bearing a region of homology to the deletion cassette in the yeast knockout strains. The resulting pool of Tamp strains were competed in batch culture. *The stationary phase experiment consisted of one growth curve and repeated sampling over 15 days while the culture remained in stationary phase (see Methods). (B) Fitness for each Tamp replicate was calculated as the slope of the change in relative frequency over time. Final Tamp fitness is determined by the median fitness of all replicates for the strain.

additional evidence that the effects of large amplifications in nutrient limitation are the result of a small number of genes, rather than the additive effect of many genes (Sunshine et al. 2015).

In this study we have expanded our focus to environmental conditions under which aneuploidy is often detrimental to growth and survival: extended stationary phase, high temperature, and treatment with the HSP90 inhibitor radicicol (Figure 2.1A). The mass overexpression of genes present on an amplified chromosome has been shown to result in stress on the cell's protein homeostasis pathways by drastically increasing the pool of proteins that

must be folded and later degraded (Torres et al. 2010, Pavelka et al. 2010, Dephoure et al. 2014). It has been hypothesized that this general stress is the basis for aneuploid cells' increased sensitivity to conditions such as high temperature and drugs inhibiting HSP90 that place additional strain on protein turnover pathways. Using the Tamp collection, we can see the degree to which these detrimental phenotypes are due to a small number of driver genes, as we predicted and observed in nutrient limitation, or due to the generalized overexpression of many genes at once.

We performed competitions of our Tamp pool in three conditions where yeast cells bearing whole chromosome amplifications proliferate more slowly or lose viability faster than euploid yeast— high temperature, prolonged culture in stationary phase, and treatment with radicicol. These competition experiments were performed in batch culture, with multiple samples taken for sequencing across approximately 17 generations of growth. Fitness measurements were calculated from the change in relative frequency of each strain across generations derived from a linear slope (Figure 2.1B, see Methods). From these experiments, we gathered fitness data from ~1500 unique Tamp karyotypes across each condition with each karyotype represented, on average, by 18 replicates from the original construction of the pool (Supplementary Figure 2.1). The fitness for each Tamp strain was determined as the median value of these replicates in each condition. Because the Tamp collection was constructed from the heterozygous deletion collection, we performed an additional competition with this pool of deletion strains and excluded any Tamp strains where the corresponding deletion strain displayed a significant fitness effect (Supplementary Table 2.3).

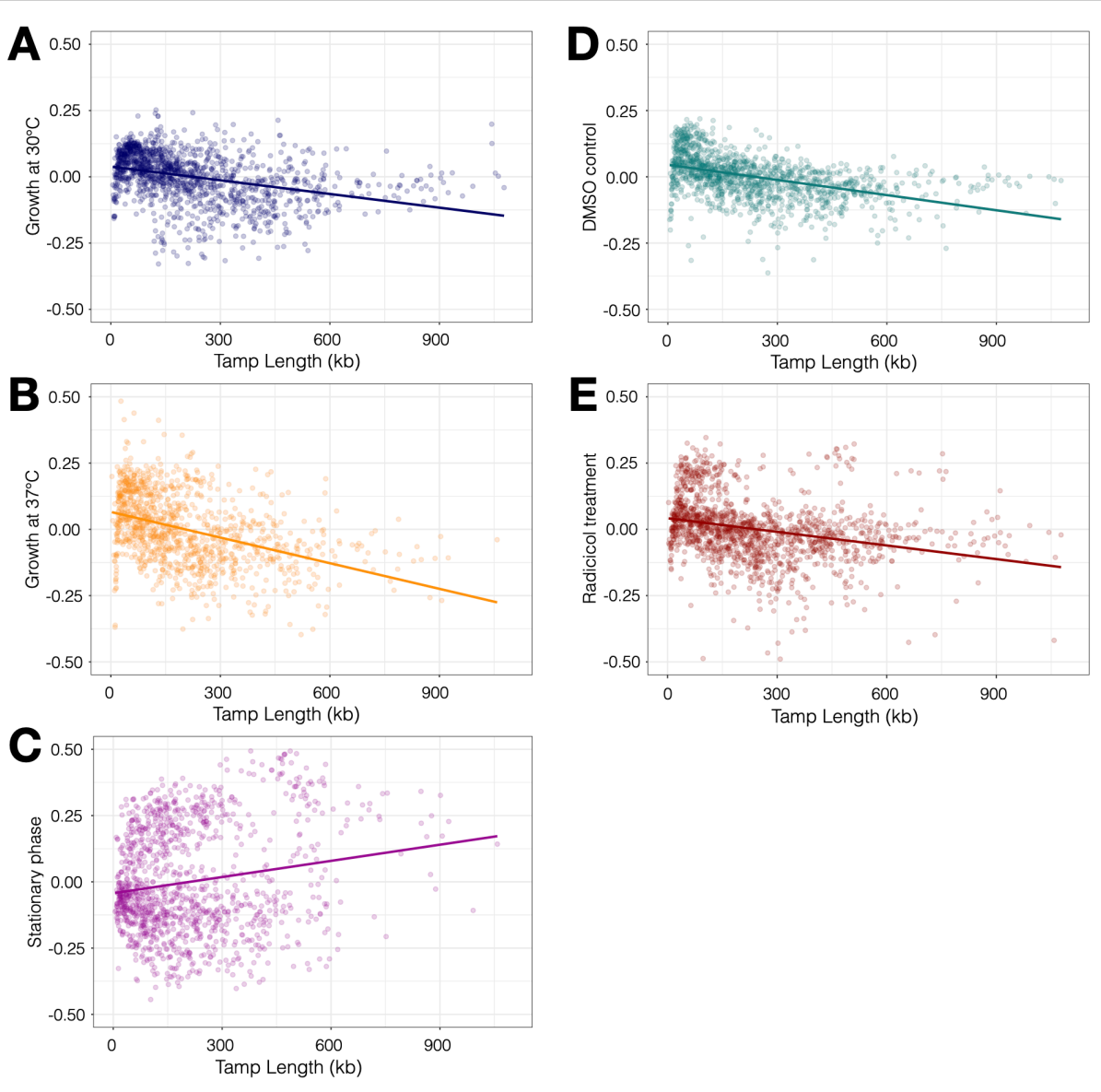


Figure 2.2: Fitness largely decreases as Tamp length increases. Fitness measurements of Tamp strains competed in (A) 30°C batch culture growth, slope = -1.72E-04; (B) 37°C batch culture growth, slope= -3.22E-04; (C) stationary phase, slope = 2.13E-04; (D) treatment with DMSO, slope = -1.88E-04; and (E) treatment with radicicol, slope = -1.72E-04.

2.2.2 Chromosomal amplifications largely decrease strain fitness

We first looked at the general trend of fitness measurements as amplification size increased (Figure 2.2). A linear trend with a decreasing slope would indicate that fitness largely

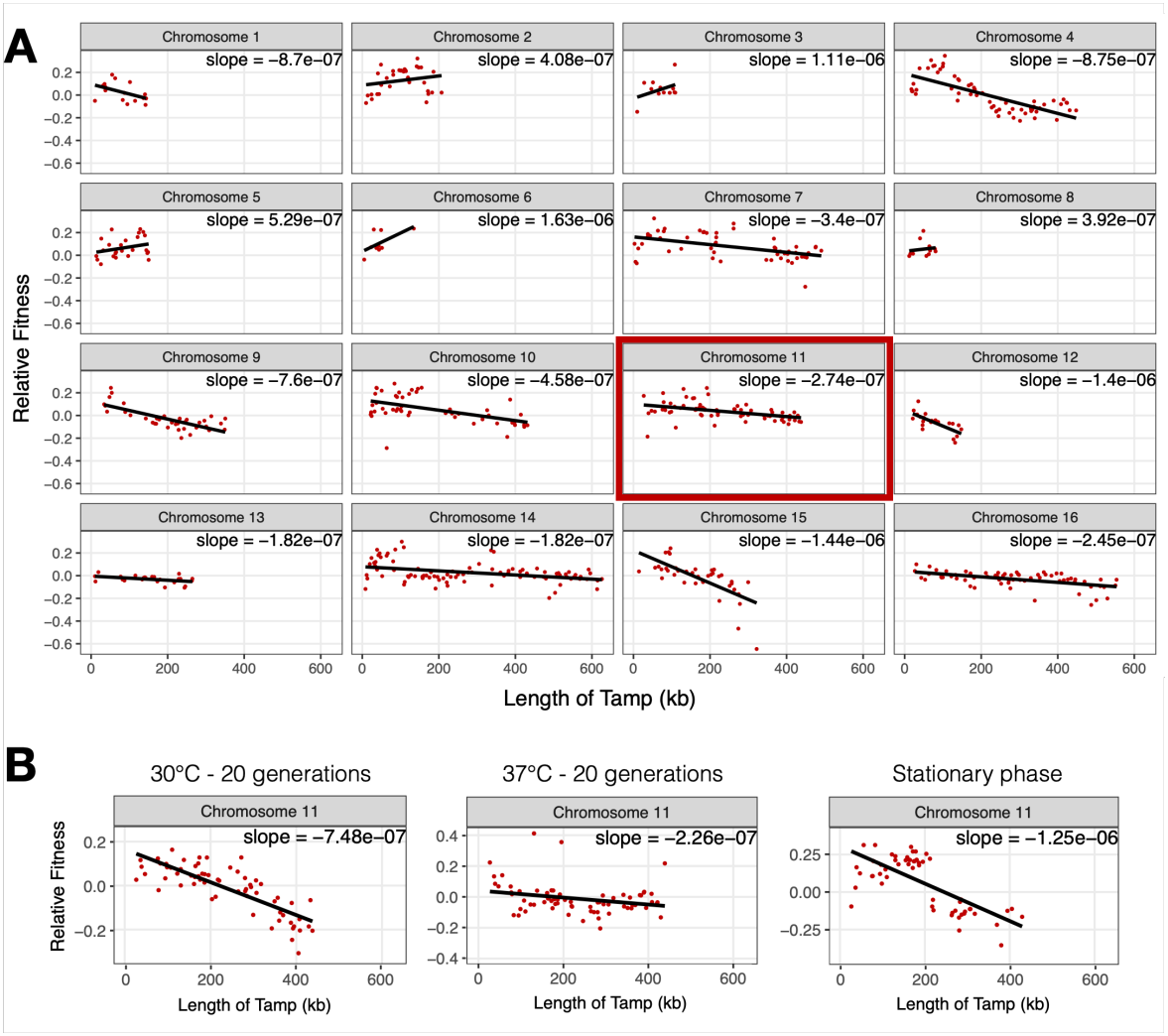


Figure 2.3: Nonlinear and condition-specific trends in fitness across chromosomes. (A) Fitness of Tamp strains grown in 20 uM radicicol by length of amplification. Data are from the amplifications of the left arm of the chromosome only. (B) Fitness of Tamps with amplifications on the left arm of chromosome 11 from competitions in low temperature, high temperature, and extended stationary phase for comparison with boxed chromosome 11 in (A).

decreases as length of the amplification increases and that the specific genes present on amplifications may matter less than the number of genes that have been amplified. Across most conditions, we found that this trend was in general correct. On average, fitness decreased by -1.28x10-4 per kb across all conditions. Interestingly, prolonged culture in stationary phase

was the only condition that resulted in an overall positive correlation between fitness and length of amplification (Figure 2.2C), while all other conditions resulted in a negative correlation. This result is consistent with the hypothesis that large amplifications largely decrease fitness, unless condition-specific beneficial genes are present on the amplification.

We then separated out chromosomes to see if more complex patterns were apparent in different regions of the genome. Because all Tamps extend from a particular region in the genome to the telomere on the same arm, we also separated our fitness data by chromosome arm to more clearly observe potential patterns. Overall, 82.5% of chromosomes displayed a negative relationship between fitness and amplification length (1, 2, 8, 0, and 3 of 16 chromosomes for 30°C, 37°C, stationary phase, DMSO, and radicicol experiments, respectively). While the overall trend remained that longer amplifications had lower fitness values, multiple regions stood out as having a nonlinear relationship between fitness and amplification length (Figure 2.3A). Additionally, these regions did not behave similarly across different environmental conditions, indicating that amplifying particular regions of the genome have contrasting effects in these different cases (Figure 2.3B). To further examine this occurrence, we plotted each Tamp's fitness by the chromosome coordinate of the amplification start site. In our previous work in nutrient limitation, this plotting method allowed us to observe a step-like pattern across chromosomes where fitness appeared to change very little until certain regions of the genome were amplified. We hypothesized that these regions contained genes that were disproportionately impactful on fitness when amplified in a particular environment. This representation of the data also seemed consistent with the hypothesis of small numbers of genes having a disproportionate effect on fitness that we previously observed in nutrient limitation, where large amplifications have been shown to be beneficial.

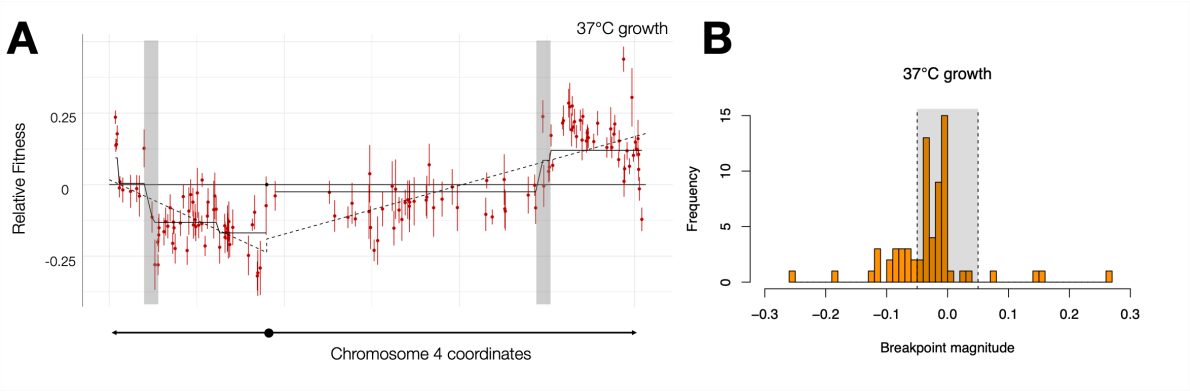


Figure 2.4: Piecewise constant model identifies candidate regions of large effect. (A) Fitness of Tamp strains with amplifications on chromosome 4 plotted by chromosome coordinate. Dotted line represents a linear regression model and solid line represents a piecewise constant model fit by 1D fused lasso. Shaded regions are breakpoints of high magnitude. (B) Histogram of breakpoints from competition in 37°C. Dotted lines represent cutoff values of +/- 0.05.

2.2.3 Identification of genomic regions with high impact on fitness

To analyze the possibility of certain genes having a disproportionate effect on fitness in more detrimental conditions, we fit a piecewise constant model using a 1D-fused lasso to our fitness data across each chromosome. In our model, we selected the largest tuning parameter within two standard errors of the best fit to limit overfitting of smaller step points resulting from experimental noise. This analysis allowed us to identify candidate breakpoints where fitness of amplifications is disproportionately impacted by a small number of genes (Figure 2.4A, Supplementary Figures 2.4-2.8). Using this method, we identified an average of 70 candidate step points that may contain genes of large effect for each experimental condition (Table 2.1). In all conditions but stationary phase, many more of the candidates were 'downsteps' - regions where fitness decreased once they were present on the amplification.

While using a relaxed tuning parameter (from the more standard of largest within one standard error of the best fit) decreased the number of small steps present in the data from an

Table 2.1: Summary of candidate breakpoints from piecewise constant models. High magnitude breakpoints are those with a magnitude greater than 0.05.

	1se model	2se model	Increased fitness	Decreased fitness	High magnitude
Low temperature	94	63	15	48	15
High temperature	101	69	8	61	23
Stationary phase	94	55	31	24	21
DMSO treatment	98	58	6	52	6
Radicicol	153	103	24	79	26
Mean	108	69.6	16.8	52.8	18.2

average of 108 steps per condition to an average of 70, we still found a high number of smaller steps in the model, indicating we may still be overfitting the data and including steps that were most likely from experimental noise. In each condition, many of the steps had a very low magnitude with a much smaller number at higher magnitudes (Figure 2.4B, Supplementary Figure 2.3). To limit our initial pool of candidate steps, we chose to focus on the step points of highest magnitude. Selecting only steps with a magnitude of 0.05 or higher narrowed our candidate pool of step points to an average starting set of 18 regions per condition (Table 2.1).

2.2.4 Effects of aneuploidy in detrimental conditions are still largely condition specific

Our work investigating the effects of large amplifications in nutrient-limited media suggests that the fitness benefits, or detriments, are largely condition-specific. To test if this result holds true in our batch culture experiments, we compared candidate step points of high magnitude to see whether they were largely shared or unique to specific growth environments. Of 91 step points across all five conditions, only 5 regions were present in multiple conditions, and none were shared between all conditions (Figure 2.5). Furthermore, all but one of these shared candidate regions were downsteps, regions that decreased fitness when amplified, while

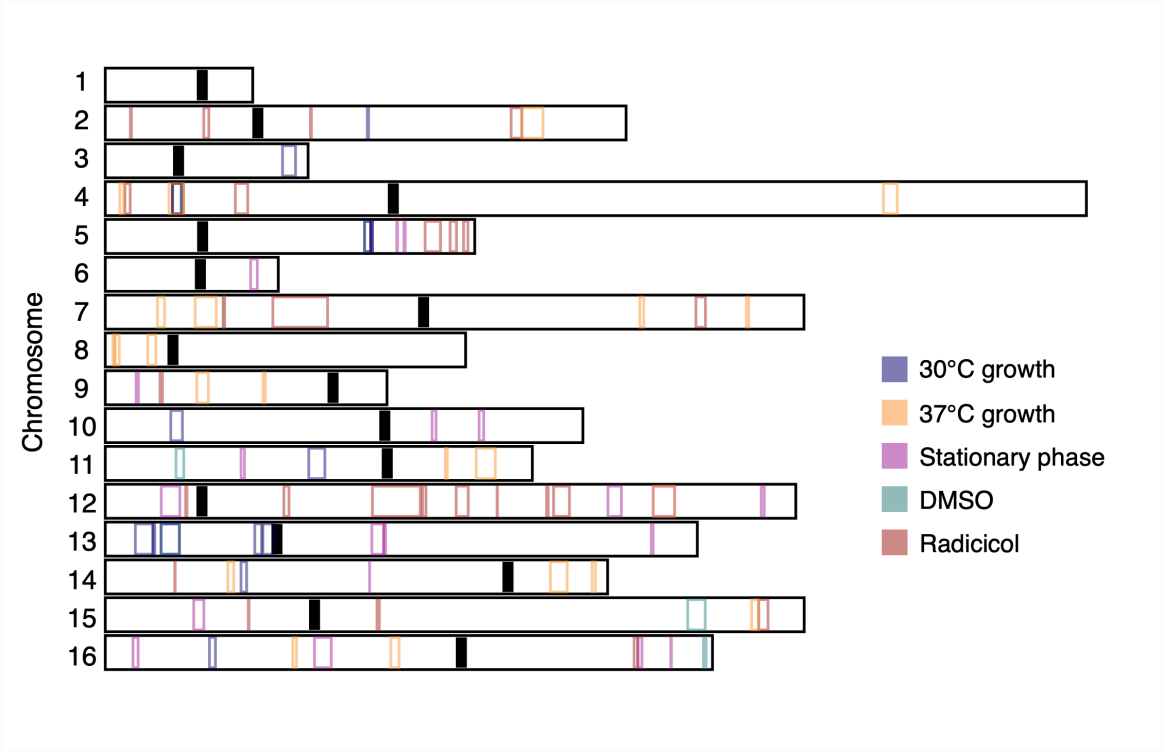


Figure 2.5: Regions of large effect identified by piecewise constant model are condition specific. Breakpoints with a magnitude greater than 0.05 as identified in piecewise constant model are outlined for each condition. Centromeres are indicated by black rectangles. Genes located withing candidate breakpoint regions are listed in Supplementary Table 2.4.

the exception was a region on chromosome 12 that was an upstep in extended stationary phase and a downstep in high temperature. This breakpoint highlights an interesting case where amplification of a gene may have contradictory effects in two conditions.

2.3 Discussion

Large chromosomal amplifications are a relatively common error in mitosis and meiosis and can cause significant problems on a cellular and organismal level. However, studies in yeast species have repeatedly highlighted the complex and sometimes contradictory effects aneuploidy can have on fitness. While large amplifications are a frequent mechanism by which organisms can adapt to new environments, they also result in sensitivity to an array of environmental

stressors. In yeast, amplifications of most whole chromosomes result in lowered tolerance of high temperatures, compounds impacting protein homeostasis, and prolonged stationary phase (Torres et al. 2007). It has been hypothesized that these environments exacerbate the stress on protein turnover pathways from expressing, folding, and degrading products of many amplified genes at once.

To investigate whether fitness deficits associated with large amplifications is dependent solely on the number of genes amplified or the specific genes present on the amplification, we used a collection of yeast with large amplifications tiled across the entire genome (Sunshine et al. 2015). Using this collection, we found that while fitness does decrease as length of the amplification increases, the correlation is weaker than would be expected if only the number of genes present on the amplification contributed to fitness. In some cases, fitness even increased with increasing Tamp length, suggesting the presence of one or more genes that provided an adaptive benefit when amplified in these conditions. By applying a piecewise constant model to our data, we were able to identify small regions of the genome that impact fitness in both positive and negative ways, indicating the negative effects of large amplifications may also be driven by individual genes in certain cases. This result is consistent with our Tamp competitions in nutrient limitation, where we found discrete regions of the genome driving fitness effects.

It is important to note two potential limitations of data presented here— errors present in our Tamp pool and limitations of our model. First, as we have previously described, larger Tamps are more likely to have errors resulting from the pooled construction of the collection, including incomplete amplification extending all the way to the telomere. We have attempted to address this problem with the inclusion of multiple replicates of each Tamp karyotype and stringent filtering steps during analysis to exclude Tamps with high variability in fitness.

However, it is possible that these errors may still be confounding our fitness measurements and models, particularly of Tamps on larger chromosome arms. The collection was also designed on the backbone of the yeast knockout collection, which has been shown to harbor mutations and chromosome abnormalities resulting from adaptation to non-essential gene deletions (Puddu et al. 2019). While this undoubtedly introduces noise into our experimental data, calling breakpoints as the result of the pattern that many individual Tamp strains together form limits the impact of these errors somewhat. Second, despite allowing for a larger tuning parameter and simpler model to limit overfitting of our data, we still found that many breakpoints called by the model were small and likely attributable to noise in the data. To account for this, we limited our focus of potential candidate genes to only those present in breakpoints of high magnitude. This narrowed scope leaves open the chance that we have filtered out true candidates whose effects were masked by experimental noise. The presence of many small breakpoints in the piecewise constant model could also be indicative of the likelihood that both length of amplification and gene content contribute to fitness. In this case, the model may result in many small breakpoints attempting to follow a linear trend that is interspersed by fewer, large breakpoints caused by large effect genes.

In this study we focus on the genes that may be impacting fitness when amplified, but other genomic features may also contribute to observed fitness effects. We have defined fitness here as a change in relative frequency over time, a metric that encompasses both the ability of cells to maintain viability and the relative speed at which they grow and divide. In order to divide, a cell must complete replication of all chromosomes. In yeast, this process initiates at autonomously replicating sequences (ARSs) and proceeds bidirectionally across chromosomes. Our Tamp collection relies on native ARSs to replicate our amplifications; however, long

stretches of a chromosome without a strong or early firing origin of replication may impact the time it takes for a cell to divide. In this case, Tamp strains with amplifications extending into long stretches devoid of an early origin would show a decreased fitness compared to a strain with a slightly longer amplification encompassing an ARS, resulting in an upstep in fitness in our assay. The impact of replication origins and other genomic features such as transposable elements requires further investigation.

The contributions that individual genes make towards overall cellular fitness of aneuploid cells have implications even beyond how microbes may adapt to new environmental challenges. As discussed earlier, aneuploidy is exceedingly present in nearly all cancers. Additionally, cancers derived from different tissue types display different recurring amplifications or deletions of specific chromosomes (Kim et al. 2013). This finding suggests that specific genomic rearrangements may provide a proliferative advantage to cancers present in different parts of the body, possibly due to the amplification or deletion of specific genes impacting malignancy. This work lends strength to the hypothesis that negative impacts of aneuploidy may also be driven by copy number changes of individual genes, bolstering the possibility of drugs and therapeutics specifically targeted to the karyotype of the tumor. Because cancer cells have unique karyotypes often with many different chromosome copy number changes, they may be uniquely sensitive to specific compounds that target amplified or deleted genes in their genome.

2.4 Materials and Method

Strains used in this study: Strains and primers used in this study are listed in Supplementary Tables 2.1 and 2.2. The genome-wide Tamp pool construction is described in (Sunshine et al. 2015). Tamp strain competitions were carried out in minimal media +20mg/L histidine with additional additives if indicated. Heterozygous deletion strain competitions were carried out in

minimal media +20mg/L histidine and +20 mg/L uracil. Yeast were grown at 30°C unless otherwise noted.

Plate reader experiments to determine radicicol concentration: The radicicol concentration used for Tamp competitions was determined by testing the effects of 0μM, 2.5μM. 5μM, 10μM, and 20μM radicicol on doubling time of wild type cells and cells bearing a large chromosomal amplification (YOR050C_Tamp, YGR238C_Tamp, YBL107C_Tamp; Supplementary Table 2.1). Log-phase cells were diluted to an OD600 of 0.05 in appropriate synthetic media in a 96-well flat-bottom plate. Data for growth curves were obtained on a BioTek Synergy H1 Multi-Mode plate reader. A concentration of 20μM radicicol resulted in greatest stratification of growth rate effects on Tamp strains with only modest effects on wild type cells (Supplementary Figure 2.2).

Competition experiments: Competitions with the Tamp pool were performed in batch culture as follows. Competitions with corresponding strains from the yeast heterozygous deletion collection (Tong et al. 2001) were conducted in parallel in B+His+Ura medium using the same experimental procedures.

Temperature competitions (30°C and 37°C): Two replicate flasks of B+His medium were inoculated with the Tamp pool for each temperature at an initial OD600 of 0.05. Cultures were grown in shaking incubators for approximately 17 generations, with back dilutions to an OD600 of 0.05 approximately every 36 hours. At each back dilution and at the conclusion of the experiment, 2 mL of culture were harvested and processed for genomic DNA using a modified Smash-and-Grab protocol (Hoffman and Winston 1987).

Stationary phase competitions: Two replicate flasks of B+His medium were inoculated with the Tamp pool at an initial OD600 of 0.05. Cultures were grown in a shaking incubator for

approximately 2 weeks. Samples were taken for sequencing on days 1, 5, 9, 11, and 13. To avoid sequencing residual DNA from dead cells, 750 µL of culture were diluted in 4 mL of fresh medium and grown for approximately 16 hrs. 2 mL of culture were then harvested and processed for gDNA.

Radicicol competition: Two replicate flasks of B+His media with either 20 µM radicicol in DMSO or DMSO only (final concentration 2mM or 0.0015% DMSO) were inoculated with the Tamp pool at an initial OD600 of 0.05. Cultures were grown in a shaking incubator for approximately 19 generations, with back dilutions to an OD600 of 0.05 approximately every 31 hours. In addition to the 2 mL samples harvested at every back dilution and at the end of the experiment, samples of at least 2 OD units of cells were harvested at hours 10.5, 14, 17, and 23.5 and processed for gDNA.

Fitness determination by Illumina barcode sequencing: Barseq libraries were prepared using primers AK1-3 for the Tamp pools and AK8-9 for the deletion pools (Supplementary Table 2.2) and purified using a Zymo Research DNA Clean & Concentrator kit (cat. no. D4004). Libraries were pooled at an equimolar ratio, size selected with Beckman Coulter Agencourt AMPure XP PCR Purification beads (item no. A63880) at a ratio of 1:1, and verified on Invitrogen Novex 6% TBE gels (cat. no. EC62655BOX), Agilent D1000 DNA ScreenTapes (cat. no.5067-5582), and/or a Roche KAPA Library Quantification Kit (kit code KK4828). Each pool was sequenced on an Illumina NextSeq, generating a total of 410,136,912 reads. Fitness calculations were made with custom python and R scripts (Github) based on analysis approaches published previously (Sunshine et al. 2015; Payen et al. 2016).

Barcode reads were filtered using the FASTQ Quality Filter from the FASTX-Toolkit to eliminate barcodes with any base Q scores of less than 20. Remaining read pairs were assigned

to a Tamp strain based on the 20-bp yeast deletion collection uptag or downtag, indicating genomic location of the *KanMX* cassette used as the Tamp initiation site, and an additional 12-bp barcode on the opposite read marking transformation replicates of each strain. Strain-replicate pairs were further filtered for sufficient number of reads with a cutoff of an average of 5 reads per time point, and if necessary, only the 100 replicates of each strain with the highest numbers of reads were used to determine fitness. Read counts were then normalized by taking the $\log_2$ ratio of the frequency of the replicate at each time point relative to the initial frequency in the pool. Read counts of zero were considered missing data. The $\log_2$ ratios were plotted against the generation time (or day number for the stationary phase experiment) yielding an average slope for each replicate. Fitness was determined as the median fitness among replicates. To account for possible incorrect karyotypes, such as truncation of the desired amplification, Tamp strains with high variability in replicate fitness values, defined as having a standard error over 1 standard error above the mean standard error, were discarded.

Identification of candidate driver genes: Competitions of the Tamp pool in glucose-, phosphate-, and sulfate-limitation previously conducted (Sunshine et al. 2015) showed a stepwise pattern in fitness effects of amplifications when plotted against the genomic location of the amplification start site, indicating that small regions of the genome may contribute disproportionately to these fitness effects. To identify these regions of large effect, we fit a piecewise constant model using a 1D fused lasso. Initially, we used 5-fold cross validation to select the largest tuning parameter that resulted in a piecewise constant model within one standard error of the best possible fit. While this method identified many large step points for further study, we found that in some cases, the model overfit the data and included step points not likely to actually include driver genes. To limit this overfitting, we relaxed the model,

allowing for the largest tuning parameter within two standard errors of the best fit. We then prioritized step points of the highest magnitude for further investigation. Candidate step regions are listed in Supplementary Table 2.4.

Chapter 3. Validation of impactful genes on Tamp fitness

The work described here was performed by Abigail Keller. The competition data from nutrient-limited chemostats used to identify candidate step points was gathered by Anna Sunshine, and published in PLoS Biology (Sunshine et al. 2015). Original analysis using DNAcopy was performed by Anna Sunshine and Kean Ming Tan. Original development of the validation strategy described here was by Anna Sunshine and Maitreya Dunham. Omar Kunjo provided assistance with deletion collection verification, tetrad dissections, and rescue by mating experiments. Noah Hanson performed Nextera Sequencing on provided samples.

The Tamp collection is a powerful tool for investigating both how large amplifications impact fitness in *Saccharomyces cerevisiae*, and how individual genes may contribute to these fitness effects. Due to the tiled nature of the collection, fitness differences between two similar Tamp strains are likely due to the presence or absence of a small genomic region on the amplification. On average, this amounts to four genes between neighboring Tamp strains; however, in practice the difference may be up to 50 genes between Tamps. While the genome-wide Tamp competitions resulted in thousands of fitness measurements across our selected conditions, further investigation is needed to determine if fitness effects of large amplifications can be attributed to a small number of genes. The fitness breakpoints described in Chapter 2 as well as Sunshine et al. 2015 provide a candidate list of genes that may be disproportionately impactful when amplified. We devised a strategy to systematically return each gene across two of these breakpoints to a euploid complement, allowing us to assay the contribution of each gene to the overall fitness of the Tamp. For this study, we focused on a region on chromosome 10 spanning one breakpoint observed in glucose-limitation and one breakpoint in phosphate-limitation.

3.1 Introduction

The large number of genes impacted by aneuploidy and large copy number variations complicates the linking of specific gene copy number alterations to phenotypes observed. In human genomic disorders this is particularly difficult because identifying causative genes may depend on multiple affected individuals with variable, but overlapping CNVs. This can be unlikely depending on the surrounding genomic architecture and whether common breakpoint sites are present in the area as well as the rarity of the CNV in the human population. Additionally, the effects of a CNV depend heavily on the genetic background it occurs in and can be confounded by other mutations and variation in the genome.

In yeast, we have several cases where the effects of large amplifications or deletions are narrowed down to copy number changes of a single gene. The yeast species *Candida albicans* is a frequent human pathogen and often develops resistance to the antifungal azole drugs used to treat infections. A study of clinical and laboratory azole resistant strains revealed a high level of aneuploidy, particularly of trisomy on chromosome 5 or an isochromosome of chromosome 5L (Selmecki et al. 2006). Further investigation into the recurrence of this particular chromosome alteration revealed that resistance was conferred by the amplification of two specific genes—*ERG11*, the drug target of fluconazole, and *TAC1,* a transcriptional regulator of the drug efflux pumps *CDR1* and *CDR2* (Selmecki et al. 2008).

We have also seen the benefits of aneuploidy in nutrient-limited chemostat evolution experiments. Under sulfate-limitation, amplifications of the entire or left arm of chromosome 2 are frequently amplified, and evolved clones bearing these amplifications outcompete euploid ancestors in sulfur-limited conditions (Gresham et al. 2008). Several lines of evidence have linked this benefit to the presence of *SUL1*, a high-affinity sulfate transporter responsible for

uptake of sulfate molecules. First, amplification of *SUL1* alone in inverted palindromic repeats has been observed in sulfate-limited chemostats and provides a similar adaptive benefit (Payen et al. 2014). Additionally, Tamp strains bearing the *SUL1* gene all display a heightened fitness in sulfate-limited growth (Sunshine et al. 2015). When single-gene deletions are paired with a Tamp strain containing *SUL1*, only deletion of *SUL1* eliminates the fitness benefit observed in the amplification (Sunshine et al. 2015).

Each candidate region identified in Chapter 2 of this thesis contains multiple genes, as well as additional genetic elements such as transposable elements, that together cause the fitness jump observed in our data. To determine whether a single gene could drive the fitness effects observed at one of these candidate sites we employed a similar method to the one used to calculate the effects of single genes on the Tamp bearing *SUL1.* For two breakpoints identified on chromosome 10 by the piecewise constant model developed in Chapter 2, one in glucose limitation and one in phosphate limitation, we verified yeast knockout collection strains with deletions within and bordering these breakpoints and mated them to a strain bearing a Tamp including the breakpoint. By returning each gene to a euploid copy number and competing the resulting pool in each nutrient-limitation, we can assess the contribution of each gene to the fitness of the amplification.

3.2 Results

3.2.1 Selection of candidate region for analysis

Work with the Tamp collection by Anna Sunshine in nutrient limitation suggested that fitness in this condition was largely determined by the presence or absence of a relatively small number of genes on the amplification (Sunshine et al. 2015). Her work yielded a total of 181 breakpoints (77 downsteps and 104 upsteps) across sulfate-, glucose-, and phosphate-limitation.

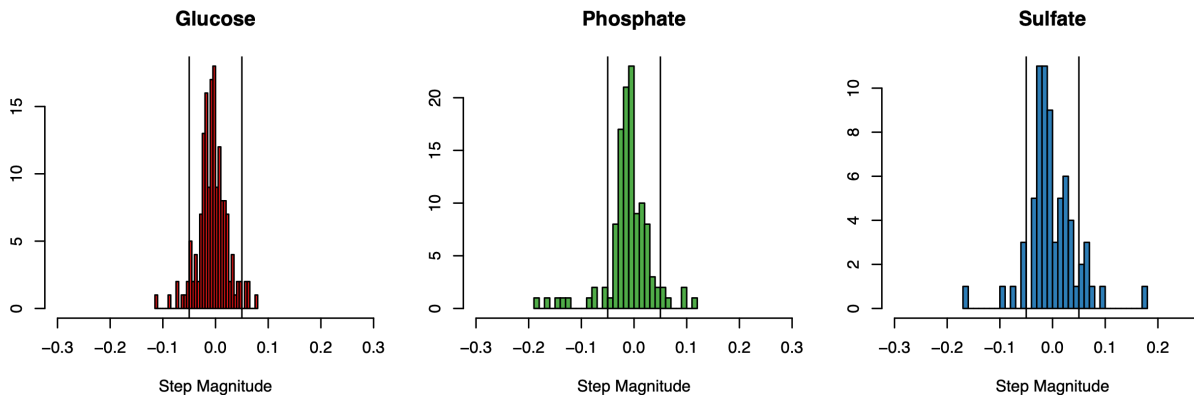


Figure 3.1: Distributions of breakpoints identified in piecewise constant model in nutrient-limited chemostat growth. Steps were identified from competition data in (A) glucose-, (B) phosphate-, and (C) sulfate-limited chemostats. Vertical lines indicate magnitude cutoff values of +/- 0.05.

While work on Tamp competitions in detrimental conditions described in Chapter 2 was still ongoing, we leveraged this complete dataset to optimize our method of calling fitness breakpoints and screening candidates for potential drivers of fitness effects on large amplifications. Applying the piecewise constant model to the nutrient limitation data resulted in a similar pattern of breakpoints to what we observed in our detrimental conditions— a large number of small magnitude step points with a smaller set of higher magnitude step points (3.1). We filtered our breakpoints for those with a magnitude of 0.05 or higher, resulting in 13, 16, and 14 candidate regions in glucose-, phosphate-, and sulfate-limitation respectively. We compared this list to the list of candidate regions identified in the original analysis using the program DNAcopy (Sunshine et al. 2015). All but four breakpoint regions from each condition overlapped with a region called in the DNAcopy analysis, raising confidence in these candidate regions for further study (Figure 3.2).

From the original analysis performed on this dataset, we knew that fitness effects observed in nutrient limitation were largely condition specific. Comparing our narrowed down

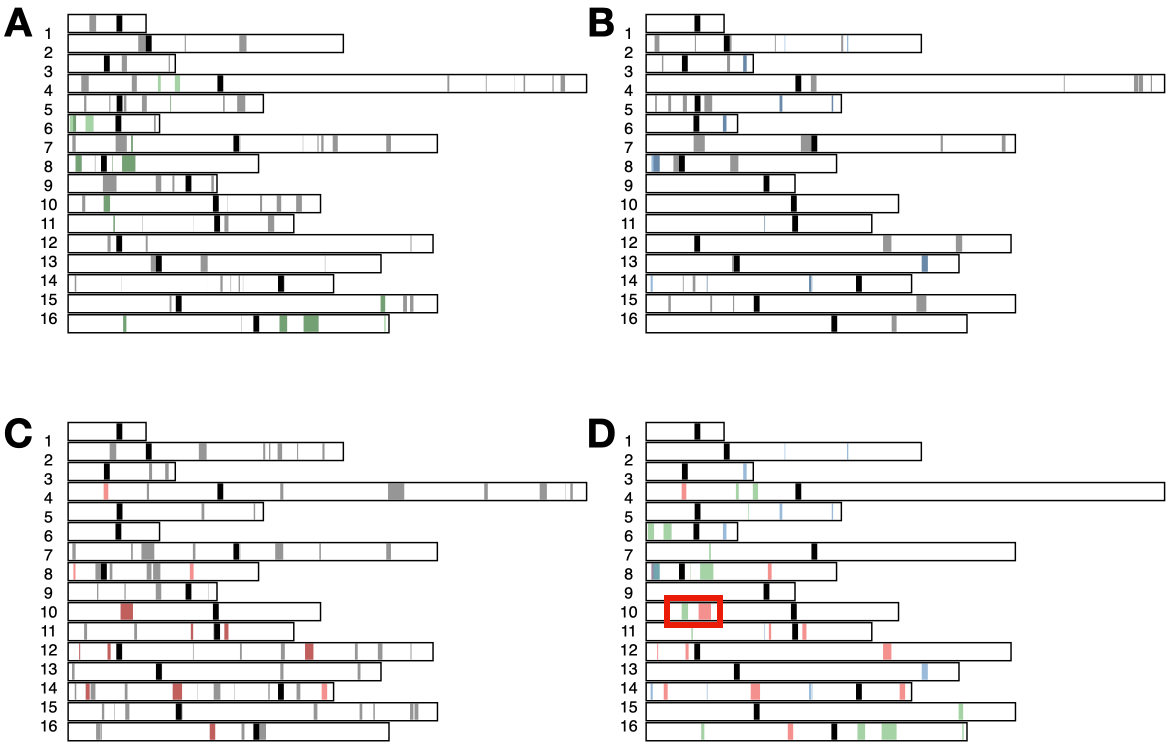


Figure 3.2: Locations of candidate breakpoints across the genome for nutrient-limited conditions. Steps were identified from competition data in (A) phosphate-, (B) sulfate-, and (C) glucose-limited chemostats. (D) Overlap of breakpoints in all nutrient-limited conditions. Steps in grey were identified using DNAcopy. Steps in green, blue, and red are were identified using the piecewise constant model in phosphate-, sulfate-, and glucose-limited chemostats respectively. Region boxed in red contains breakpoints investigated in this study.

list of candidate breakpoints across conditions we found that only one region on chromosome 8 had overlapping breakpoints from all three conditions (Figure 3.2D). Apart from this region, no other candidate breakpoints were shared between datasets. We ultimately selected the left arm of chromosome 10 to optimize our validation strategy (Figure 3.2D, boxed area). Competitions in both glucose- and phosphate-limitation resulted in a fitness downstep on this chromosome arm, though at two nearby but non-overlapping regions (Figure 3.3). This allowed us to test two

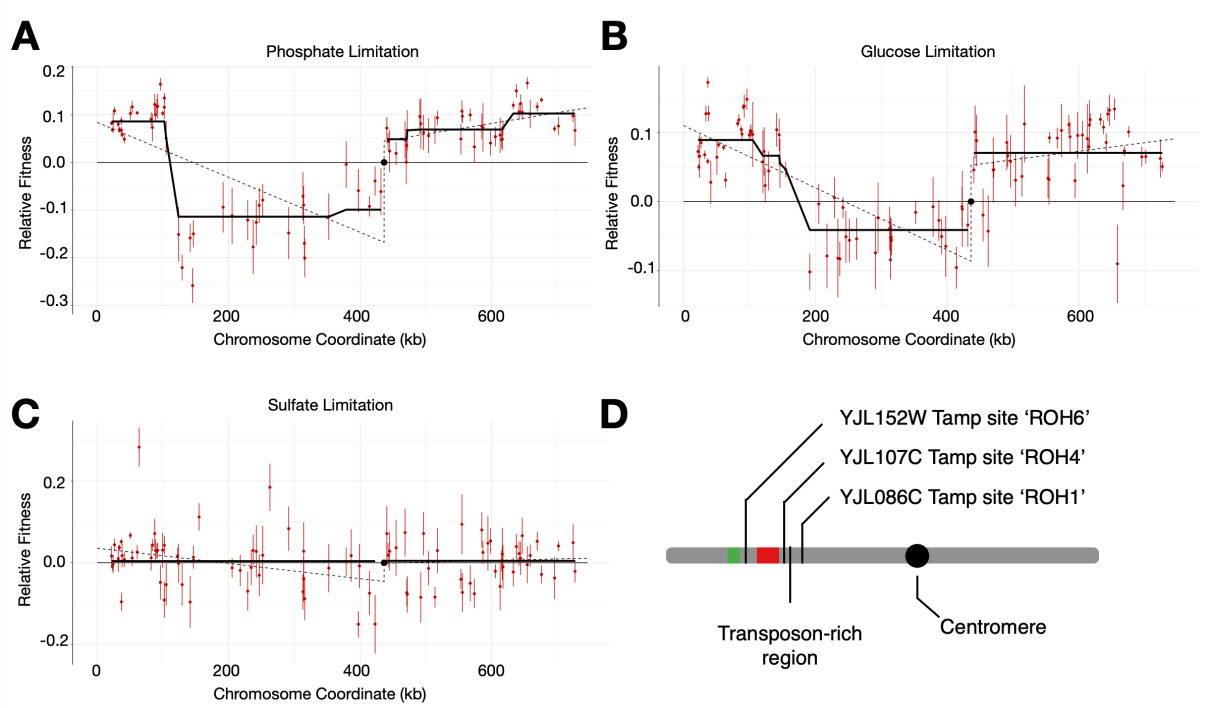


Figure 3.3: Fitness of Tamps with amplifications of chromosome 10 in nutrient-limited chemostats. Data shown from (A) phosphate-, (B) glucose-, and (C) sulfate-limited chemostats. Dotted lines represent a linear regression model and solid lines represent a piecewise-constant model fit by the 1D-fused lasso. (D) Schematic of chromosome 10 with the phosphate-limited breakpoint in green, glucose-limited breakpoint in red, and locations of the independently generated Tamp strains labeled.

breakpoints simultaneously while using sulfate-limitation, where no fitness breakpoints were observed on this chromosome arm, as a control.

3.2.2 Generation of individual large Tamp strains

The first component needed to assay our candidate breakpoint was a haploid strain bearing a single long amplification that spanned the breakpoint and extended to the left telomere of the chromosome. I selected three open reading frames (ORFs) to serve as target regions of homology for the chromosome fragmentation vectors to generate three separate Tamp strains, two with amplifications spanning both breakpoints and one with an amplification initiating in between the two breakpoint regions (Figure 3.3). All three target sites are either dubious or

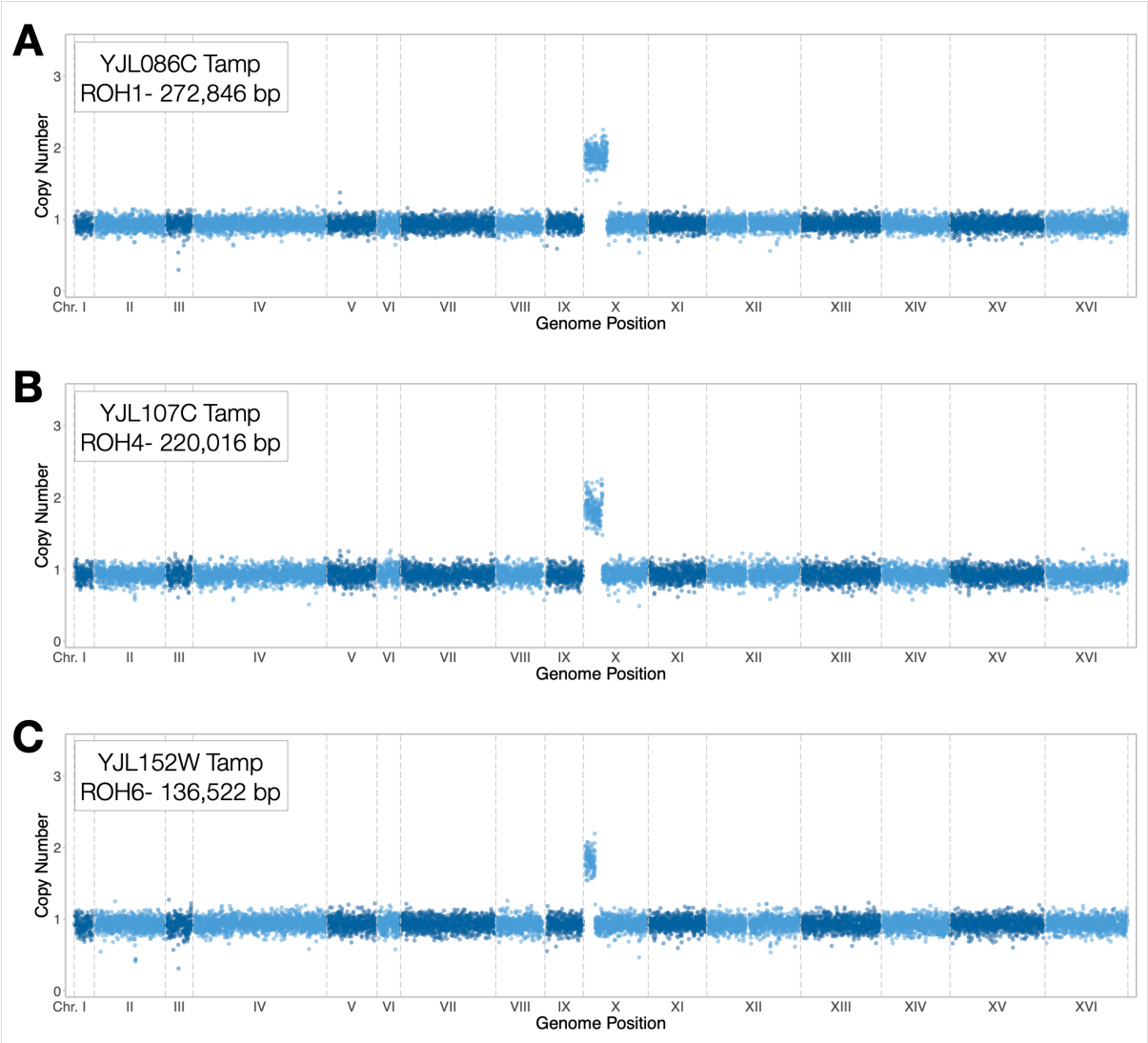


Figure 3.4: Copy number plots from whole-genome sequencing of independently generated Tamp strains. (A) Tamp with 272,846 bp amplification starting at YJL086C. (B) Tamp with 220,016 bp amplification starting at YJL107C. (C) Tamp with 136,522 bp amplification starting at YJL152W. All three strains were verified to begin at the correct site and extend fully to the telomere. Data points represent copy number based on 1000bp sliding window.

uncharacterized ORFs to limit the possible effects of amplifying a partial copy of a gene. The area directly upstream of the glucose breakpoint contained a transposon rich region, so I chose the uncharacterized ORF YJL107C just outside of this area and the dubious ORF YJL086C further upstream. The dubious ORF YJL152W, located between the glucose and phosphate

breakpoints, was selected to provide an additional, more targeted look at the phosphate breakpoint without any confounding effects from the larger amplification.

Small ~300 bp regions of each ORF were amplified and cloned into a chromosome fragmentation vector (CFV) (strain YMD3004) used to generate Tamp strains (see Methods). The telomere seeding sequence and centromere present on the vector cause the DNA to act as a small chromosome with a double-stranded break when linearized prior to transformation of a yeast cell. The amplicon serves as a region of homology (ROH) to allow the cell to repair off of the native chromosome, resulting in a duplication of the region from the targeted site to the telomere, present on an independent chromosome. Haploid yeast were transformed with CFVs made from each of the three target sites. All three CFVs successfully generated a Tamp strain at the targeted site that fully extended to the telomere and contained no other large structural abnormalities (Figure 3.4). These strains were then mated to deletion collection strains spanning the candidate breakpoints to systematically return each gene to a euploid complement in the context of a large amplification.

3.2.3 Rescue by mating strategy for obtaining deletion strains of essential genes

Strains bearing deletions from the candidate breakpoint regions were obtained from the haploid yeast knockout collection (Tong et al. 2001). This collection was derived from the diploid heterozygous knockout collection by sporulation and selection for the deletion *KanMX* marker. Strains with a deletion in an essential yeast gene, however, would not survive this process. To avoid individually transforming each heterozygous deletion with the Tamp CFV and potentially introducing additional structural abnormalities, we optimized a ‘rescue by mating’ strategy (Haarer et al. 2011). This strategy was developed in a screen for complex haploinsufficient interactions, a variation on a synthetic genetic array (SGA) approach looking

for negative phenotypic consequences of two heterozygous deletions. This approach leverages the discovery that ultimately inviable spores are capable of germinating and mating prior to death, likely because maternal factors may linger long enough to complete one or more cell cycles (Taxis et al. 2005). By mating to a complementary partner immediately following spore dissection, normally lethal haploid genotypes can still be used for diploid strain construction.

For the candidate breakpoints on chromosome 10, we compiled a list of deletion collection strains with gene knockouts spanning each breakpoint, including approximately five genes on either side of the breakpoint (Supplementary Tables 3.1 and 3.2). This resulted in 23 strains for the phosphate breakpoint and 33 strains for the glucose breakpoint. When possible, we included strains with deletions in dubious or uncharacterized ORFs to account for the possibility that ORFs with an unknown function may have a fitness effect when present in an altered copy number in the context of a large chromosomal amplification. We first attempted to gather all strains from the haploid deletion collection and verify barcodes before mating them to our haploid Tamp strains. From both breakpoints, we were able to verify barcodes for 50 out of 56 of our desired strains, either directly from the haploid deletion collection, or by re-sporulating from the heterozygous diploid collection. We were ultimately unable to confirm the correct barcode in two deletion strains, the uncharacterized ORF YJL163C and the cAMP-dependent protein kinase catalytic subunit encoding gene *TPK1*, possibly due to an underlying contamination issue in the collection. Because *TPK1* was a strong candidate gene that may be driving the fitness effects across the phosphate breakpoint, we decided to reconstruct this strain for our pool (see Methods).

To obtain the remaining gene knockouts, we confirmed barcodes from the heterozygous diploid collection and proceeded to our ‘rescue by mating’ strategy. The yeast knockout collection uses a *KanMX* cassette as a selection marker for the deletion. Similarly, the haploid

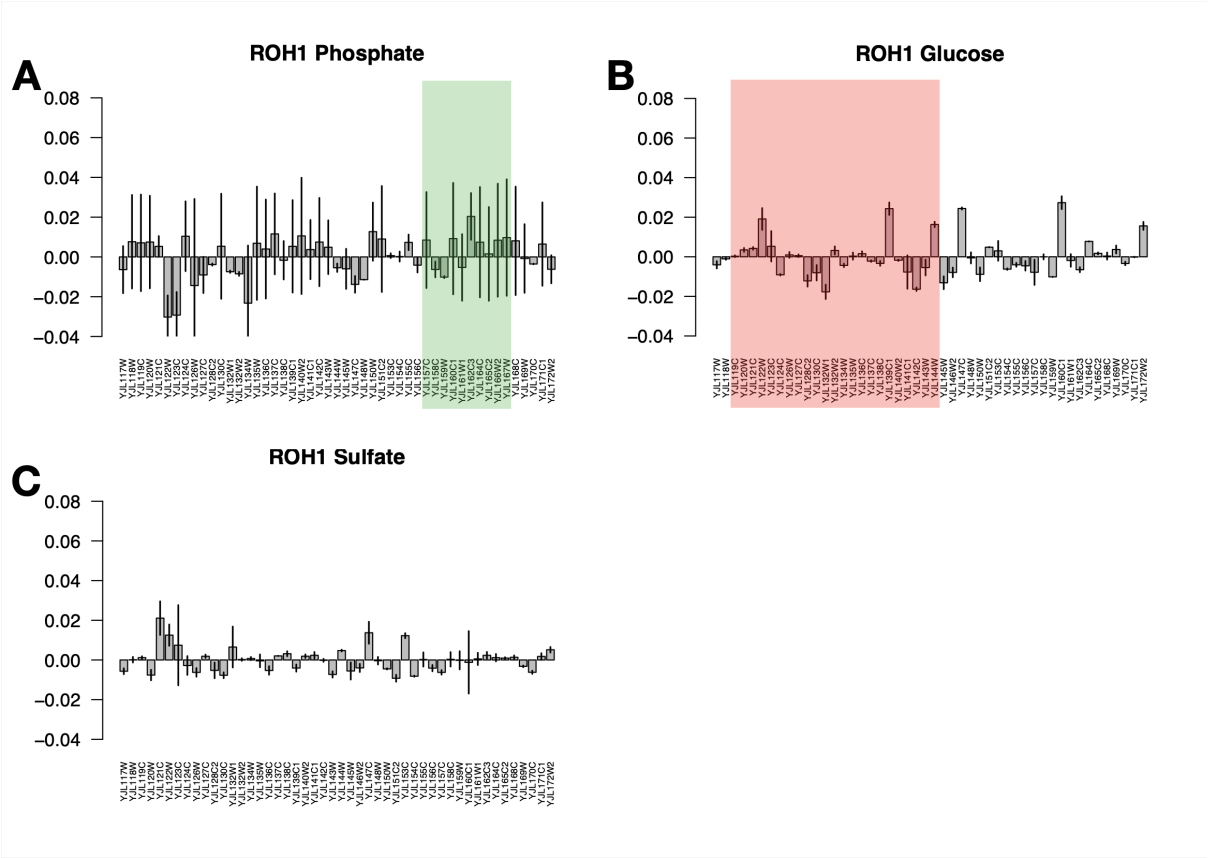


Figure 3.5: Contribution of individual genes to fitness breakpoints observed in nutrient limitation. Fitness as calculated by the change in relative frequency over time of strains with a 272,846 bp amplification on Chromosome 10L in (A) phosphate-limitation, (B) Glucose-limitation, and (C) Sulfate-limitation. Regions shaded represent genes present across the breakpoint identified in phosphate-limitation (green) and glucose-limitation (red). Error bars are +/- one standard deviation between replicate chemostats.

Tamp strains I constructed each have a *URA3* marker to maintain selection for the large amplified chromosome. After sporulation of the deletion strain, spores were digested and dispersed by sonication as part of the random spore analysis protocol (Treco and Winston 2008). These spores were then mixed with log phase cells of the haploid Tamp strain. After a period of time allowing for mating, the cells were plated on C-ura +G418 media to select for markers from both parent strains. Because neither parent could survive on this media, only successfully mated

strains with both the deletion and Tamp were able to grow. From this strategy we were able to obtain the final five strains needed for our pools.

3.2.4 Examination of candidate genes on chromosome 10 in nutrient limitation

Once all gene deletion strains had been mated to our Tamp strains, I combined them into three pools, one for each Tamp strain, for competitions in glucose-, phosphate-, and sulfate-limited miniature chemostats (Miller et al. 2013). Each pool was inoculated into two chemostats in each nutrient limitation, for a total of 18 vessels, and grown for an average of 25 generations. Samples were taken for sequencing approximately every 3 generations, allowing me to track the frequency of individual strains in the pool by their deletion barcode. As in the genome-wide Tamp pool, fitness was calculated as the slope of the $\log_2$ ratio of strain frequencies at each timepoint to the initial timepoint and normalized to the mean fitness of the pool. Overall, I was able to detect barcodes from 54 of the 56 deletion barcodes in the pool, though filtering for low read count resulted in fitness data for a total of 47 barcoded deletions.

Because both candidate breakpoints being tested span areas that appear to decrease fitness when amplified, we expect that deletion of a driver gene, thereby returning it to a euploid copy number, will increase fitness of the strain bearing the amplification. Overall, effects of single gene deletions on Tamp fitness were quite low, with no strain exhibiting an effect of more than 0.05. This was true across all three Tamp pools and all three nutrient-limitations (Figure 3.5, Supplemental Figures 3.1 and 3.2). Interestingly, competition in phosphate-limitation produced highly variable results between replicates, particularly in the two longer Tamp pools initiating at ROH1 and ROH4. Though error rates were decreased in the shorter Tamp karyotype, effect sizes were quite small. One gene deletion, *JJJ2* (YJL162C) encoding a protein of unknown function,

appeared to increase fitness of the Tamp strain; however, deletion of this gene had a similar effect in all three conditions and was not dependent on phosphate-limitation.

Competition in glucose-limitation produced similarly confusing results. The amplifications in Tamps ROH1 and ROH4 both encompass the breakpoint seen in our Tamp screen in glucose-limited chemostats (Figure 3.5, Supplementary Figure 3.1). We would expect fitness patterns across this breakpoint to be consistent in these two pools, despite a roughly 50 kb difference in size. Of all gene deletions that increased fitness of the Tamp in these two pools, only one was shared between both. This gene, *ALB1* (YJL122W), encodes a shuttling pre-ribosomal factor associated with the 60S ribosomal subunit (Lebreton et al. 2006). The small effect of this deletion in combination with the Tamp amplifications coupled with the lack of functional characterizations that may influence fitness in glucose-limitation make this an unlikely driver of fitness effects.

3.3 Discussion

Previous analysis of the Tamp collection suggested that the fitness effects of large amplifications are driven by copy number changes of a small number of genes (Sunshine et al. 2015). This was shown to be the case for amplifications of the right arm of chromosome 2 including the high-affinity sulfate transporter *SUL1.* When *SUL1* was either no longer present on the Tamp or when the Tamp was paired with a deletion of *SUL1*, the fitness benefit of the amplification was eliminated (Sunshine et al. 2015). In this work, I sought to replicate this finding in two breakpoints observed on the right arm of chromosome 10, one in glucose-limitation and one in phosphate-limitation.

The breakpoint in phosphate-limitation offered an enticing potential candidate gene of large effect in *TPK1*, a cAMP-dependent protein kinase catalytic subunit. Introduction of an

extra copy of *TPK1* was previously shown to decrease fitness in phosphate-limited chemostats and it would follow that presence of this gene on a large amplification would similarly decrease fitness (Payen et al. 2016). While the breakpoint in glucose-limitation did contain the gene *PBS2* encoding a MAP kinase kinase that, when increased in copy number, was shown to reduce fitness, it was observed to decrease fitness in all three nutrient limitations (Payen et al. 2016). For this particular breakpoint, I had two main hypotheses regarding why amplifying this region of the genome had a large impact on fitness. First, not all genes present at this breakpoint had information regarding the fitness effects of single gene copy number changes in glucose limitation. If the information gap for a particular gene was due to the highly deleterious effect of amplification, the strain may not have survived the pooled assay long enough to be detected and measured for fitness effects. Second, it is possible that fitness effects of copy number changes of particular genes only are apparent when paired with a large amplification. This could be due to synergistic effects either between the gene in question and another gene present on the amplification, or between the specific gene and simply the state of being aneuploid. Aneuploid cells experience particular stress responses (discussed further in Chapter 4) that may lead to synthetic interactions that previously have not been explored.

Unfortunately, the results of this experiment were not as clear cut as we might have anticipated. The effect sizes of returning each gene to a euploid complement were quite small, and no obvious candidates emerged as drivers of the fitness breakpoints we observed in the Tamp collection. Several possibilities may have accounted for this failure. It may simply be that more than one gene contributes to the observed fitness jump. The test case for this method of validating genes of large effect was the breakpoint encompassing *SUL1.* This breakpoint had two major factors contributing to the success of the validation: (1) Amplification of *SUL1* has repeatedly been shown to confer a huge and well characterized fitness advantage in sulfate

limitation, and (2) the number of genes present across the breakpoint was relatively small. Both breakpoints in this study are rather large in terms of number of genes, 11 in phosphate limitation and 26 in glucose limitation. It is possible that multiple genes together account for the fitness effects we have seen in these conditions.

Another possible explanation is that errors in the underlying deletion collection are confounding these results. The yeast knockout collection is a powerful and convenient tool for large screens of many individual gene deletions. But more recent sequencing of the collection has found that the yeast have adapted to life with specific gene deletions via other mutations and chromosomal rearrangements (Puddu et al. 2019). We attempted to be cautious in verifying that the strains we used were the intended deletion by sequencing the deletion barcode, but any other genomic changes were missed by this approach. The noise present in the initial Tamp screens likely was due in part to these errors, and in the smaller, more targeted approach presented here, they may have been even more impactful. We encountered an example of the effects of compensatory mutations while verifying barcodes. Deletion of *SSY5* has been shown to be lethal, though one study deems this an 'essential but evolvable' gene in that the lethality may be overcome by additional mutations. When collecting strains, we found that the *SSY5* deletion strain was viable as a haploid. Because this was likely due to other mutations that we did not want in our final pool, we decided to use our rescue-by-mating approach with the heterozygous diploid to obtain our final strain.

Overall, while we were disappointed by the lack of a clear result in this validation study, there is still room for improvement. Future attempts may require further verification of background mutations in deletion collection strains. Optimization of the rescue-by-mating protocol will certainly prove to be a critical to this effort, as the heterozygous diploid strains, with one functioning copy of the gene of interest, will undoubtedly be less prone to

compensatory mutations and rapid mating of newly germinated spores will reduce the possibility of new mutations arising. We can also select breakpoints for further study that span smaller regions of a chromosome. Because the magnitude of effect of our candidate breakpoints was large, we thought it more likely that genes of large effect were driving fitness effects. However, our results leave open the possibility that a larger than average gap between neighboring Tamps masked a more gradual decrease in fitness as more genes were present on the amplification.

3.4 Materials and Methods

Yeast media and culturing techniques: Yeast were grown at 30°C unless otherwise noted. Nutrient-limited chemostat media recipes are available at http://dunham.gs.washington.edu/protocols.shtml. Sporulation media consists of 1g/L potassium acetate, 0.1g/L yeast extract, and 0.05g/L dextrose.

Construction of individual Tamp strains: Three genomic regions were selected as Tamp initiation sites by proximity to the candidate breakpoint and classification of the ORF (dubious or uncharacterized). For each ORF, ~350 bp regions were amplified using primers OAK-033, OAK-034, OAK-038, OAK-039, OAK-041, and OAK-042 (Supplemental Table 3.3). For each pair, the forward primer contained a telomere seeding sequence and a SnaBI restriction site for linearizing the plasmid. Each amplicon was digested with AatII and BamHI and cloned into the corresponding site on the digested CFV YMD3004 to serve as a region of homology for transformation. The CFV was linearized via digestion with SnaBI and strain YAK-099 was transformed with the linear product. Final Tamp strains were validated by whole genome sequencing using custom scripts written by Chris Large (as described in Smukowski Heil et al. 2019) and coordinates and length of amplification were visually confirmed using Integrative Genomics Viewer (Robinson et al. 2011).

Verification of deletion barcodes: Strains with deletions across both breakpoints were isolated first from the haploid yeast knockout collection (Tong et al. 2001) and verified with colony PCR of the deletion "Uptag" barcode. A small amount of cells were suspended in 10uL 0.2 N NaOH, boiled for 15 minutes at 105°C, then frozen and thawed. The sample was diluted with 90uL H_2O and spun to pellet cell debris. 1uL of sample was used as template for PCR using primers OAK-052 and OAK-053 and sent to Eurofins Scientific for sanger sequencing with primer OAK-049.

Strains that either could not be recovered from the haploid collection or had an incorrect barcode were then isolated from the heterozygous diploid collection. Barcodes were amplified and verified in the same manner as for the haploid strains. If the barcode was confirmed, the strain was sporulated, digested, and dissected for tetrads. Haploid deletion strains were selected by replica plating to YPD+G418 media.

Matings and random spore analysis: 75uL of log-phase cells of each verified deletion strain were mixed with 75uL log-phase cells of each of the independently generated Tamp strains in a 96-well plate. The mixture was spun down and resuspended in 75uL YPD for mating. Samples from individual wells were randomly and periodically observed under a dissection microscope for presence of schmooing or mated cells. 10uL of each mating mixture were plated on C-ura +G418 media to select for diploid cells with both the deletion and the amplification.

For strains with deletions of essential genes, we used a random spore analysis protocol (Treco and Winston 2008) to digest and disrupt tetrads into individual spores *en masse*. Spores were incubated in 5mL water, with 100uL 100U/mL Yeast Lytic Enzyme (VWR cat. no. IC360951), and 10uL β-mercaptoethanol overnight with gentle shaking at 30°C. After adding 5mL 1.5% Triton X-100, spores were sonicated with a microtip attachment at 50% power for cycles of 30 seconds followed by 2 minutes on ice. Every three cycles, spores were checked for

dissociation on a dissection microscope. Once fully separated, spores were mixed with haploid Tamp strains in YPD and incubated overnight with gentle shaking at 30°C. The mated culture was then plated on C-ura+G418 media to select for diploids with both the deletion and the Tamp amplification.

Chemostat competition: Miniature chemostats were inoculated with the validation Tamp pools and competition experiments were run in glucose-, sulfate-, and phosphate-limited chemostat media as has been previously described (Miller et al. 2013). Each of the three Tamp pools was run in duplicate in each type of medium, for a total of 18 miniature chemostats. Samples were taken twice per day to calculate the dilution rate and OD of the cultures and prepare DNA samples to track strain frequencies. DNA was obtained using a modified Smash-and-Grab protocol (Hoffman and Winston 1987).

Calculation of strain fitness: Yeast knockout collection barcodes were amplified from the competition DNA samples using primers AK8 and AK9 and purified using a Zymo Research DNA Clean & Concentrator kit (cat. no. D4004). Libraries were pooled at an equimolar ratio, size selected with Beckman Coulter Agencourt AMPure XP PCR Purification beads (item no. A63880) at a ratio of 1:1, and verified on Invitrogen Novex 6% TBE gels (cat. no. EC62655BOX).

Fitness calculations were made with custom python and R scripts (Github). Reads were filtered using the FASTQ Quality Filter from the FASTX-Toolkit to eliminate barcodes with any base Q scores of less than 20, resulting in 6,353,073 reads across all samples. Strains were identified using the 20 bp 'Uptag' barcode of the yeast knockout collection. For each timepoint, the relative frequency of each strain was determined by dividing read counts for that sample by the total number of reads in that timepoint. Fitness was calculated as the slope of the $\log_2$ ratio of

each timepoint to T3, when cultures reached steady state, versus time in generations. This fitness was normalized to the mean fitness of the pool.

Chapter 4. Translational capacity of aneuploid yeast in phosphate-limited chemostat culture

This work was done in collaboration with Angelika Amon's lab at MIT and results from this work are published under the title "The environmental stress response causes ribosome loss in aneuploid yeast cells" in PNAS. Aneuploid yeast strains were obtained from the lab of Rong Li at Johns Hopkins. Chemostat experiments were optimized and performed by Abigail Keller with assistance from Emily Mitchell. Cells were harvested for ribosome, DNA, and RNA extractions by Abigail Keller and sent to the Amon Lab for analysis. Processing of cells for RNA, DNA, and ribosome content was performed by Arzu Sandicki and Allegra Terhorst, and the description of methods for this work is published in Terhorst et al. 2020. Conception and analysis of the experiments was by Arzu Sandicki, Allegra Terhorst, and Angelika Amon.

Aneuploidy has long been associated with negative phenotypes in both single- and multi-celled organisms. These phenotypes include sensitivity to certain environmental conditions, as discussed in Chapter 2, as well as widespread metabolic changes and growth defects when compared to euploid controls. Recent work has highlighted a specific transcriptional profile shared among aneuploid yeast cells known as the environmental stress response, that is independent of which chromosomes are altered in copy number. The presence of ribosomal proteins among downregulated genes in aneuploid cells led us to question how growth rate, degree of aneuploidy, and ribosome content are interconnected. To test these links, we grew a set of complex aneuploid strains in batch culture and in phosphate-limited chemostats to equalize growth rate among all aneuploid and euploid strains. After optimizing our chemostat protocols to allow for growth of particularly slow-growing aneuploid strains, we found that the ESR was largely dependent on proliferation rate. Furthermore, our findings indicate that the loss of

ribosomes previously observed in aneuploid strains is a result of mounting the environmental stress response to protect the cell from cellular stresses and prevent excessive enlargement of the cell during their extended cell cycle.

4.1 Introduction

The negative impacts aneuploidy has on individual cells has been widely studied. In most organisms, including *Schizosaccharomyces pombe, Drosophila* species, and humans, studies have shown that aneuploidy results in proliferative delay (Niwa et al. 2006, Lindsley et al. 1972, Segal and McCoy 1974). While copy number alterations in single-celled organisms can be a route to adaptation when environmental conditions become challenging, in less stressful conditions aneuploidy is usually poorly tolerated. Studies in disomic yeast, haploid *Saccharomyces cerevisiae* cells with a single extra chromosome, have revealed a variety of cellular stresses associated with aneuploidy (Torres et al. 2007). Analysis of these strains revealed that transcription levels of genes largely correlate with copy number, and at least some of the amplified genes showed an increased protein amount. If all amplified transcripts are also translated, this places additional stress on protein synthesis pathways, cellular chaperones, and protein degradation pathways (Torres et al. 2007). The sensitivity of disomic strains to geldanamycin, an inhibitor of the protein chaperone Hsp90, in this study was further explored in Chapter 2 with treatment of the Tamp strains with a similar compound, radicicol. Beyond protein homeostasis, disomes also show metabolic changes relative to euploid cells. Disomes have an increased uptake of glucose, resulting in a lower OD at saturation (Torres et al. 2007). Because a high proportion of cellular energy is used for protein production, and the magnitude of these effects are somewhat correlated with the number of genes present on the amplification, this

observation could also be due to the increased transcription and translation of many additional genes.

Aneuploid cells have been shown to exhibit particular transcriptional responses, known as the environmental stress response (ESR), that are consistent across chromosomes and species (Sheltzer et al. 2012). The ESR can be induced by a wide array of environmental stressors such as heat shock or nutrient limitation (Gasch et al. 2000) and results in the upregulation of ~300 genes and the downregulation of ~600 genes. This response is characterized by the coordinated upregulation of chaperones, amino acid transporters, and proteins with roles in endocytosis and proteasome activity, as well as the downregulation of transcription and translation machinery, including ribosomal proteins.

Downregulation of ribosomal proteins in aneuploid cells is particularly interesting because ribosome content is tightly tied to growth rate in cells (Schaechter et al. 1958, Ecker and Schaechter 1963, Kafri et al. 2016). Consistent with this observation, the strength of the environmental stress response is also correlated with growth rate (Sheltzer et al. 2012). To untangle the effects of aneuploidy and growth rate on ribosome content in yeast, we grew strains with multiple whole chromosome aneuploidies in phosphate-limited chemostats. In chemostats, growth rate is determined by the dilution rate of the culture due to a limiting nutrient, in this case phosphate. When growth rate was held constant for euploid and aneuploid cells, the ESR normally caused by aneuploidy disappeared, indicating that the strength of the ESR is determined by proliferation rate of cells. Additionally, analysis of ribosomal fraction in this context revealed that the reduction in ribosome content in aneuploid cells relative to euploid cells observed in batch culture growth was no longer evident when growth rate is controlled (Terhorst et al. 2020). This suggests that induction of ESR in aneuploid cells triggers the loss of ribosomes

and results in the decreased cellular density that has been observed in aneuploid cells (Altenburg et al. 2019, Gasch et al. 2000).

4.2 Results

4.2.1 Growth of auxotrophic yeast strains in phosphate-limited chemostats requires additional amounts of supplemental nutrients than batch culture.

To begin addressing ribosome content and translational capacity of aneuploid cells, I first had to adapt our chemostat protocols for the growth of auxotrophic strains. We took advantage of a previously developed set of isogenic, stable, aneuploid yeast with variable numbers of chromosomes generated through sporulation of triploid and pentaploid strains (Pavelka et al. 2010, Supplementary Table 4.1). These strains ranged from a near haploid to near triploid complement of chromosomes. Normally, cells grown in the chemostat are prototrophic for simplicity. In the past, we have grown the Tamp collection with supplemented histidine and the yeast knockout collection with supplemented histidine and uracil at standard levels of 20mg/L each without issue (Sunshine et al. 2015, Payen et al. 2016). While the aneuploid strains were of the S288C background, like both collections we had successfully cultured in the chemostat, they also had four auxotrophies, being deficient in uracil, histidine, leucine, and tryptophan production (Pavelka er al. 2010).

I used the haploid strain isogenic to the aneuploid strains to test whether growth in the chemostat was limited by phosphate, as intended, or by the supplemental nutrients required for the strains' auxotrophies. If the supplemental nutrients were not present in excess, one of these compounds could be limiting growth resulting in confounding variables. I cultured the haploid test strain in standard amounts of additives (1X = 20mg/L uracil, 20mg/L L-histidine HCl, 20mg/L L-tryptophan, and 1g L-leucine) as well as 0.5X and 2X amounts. Within 40 hours of

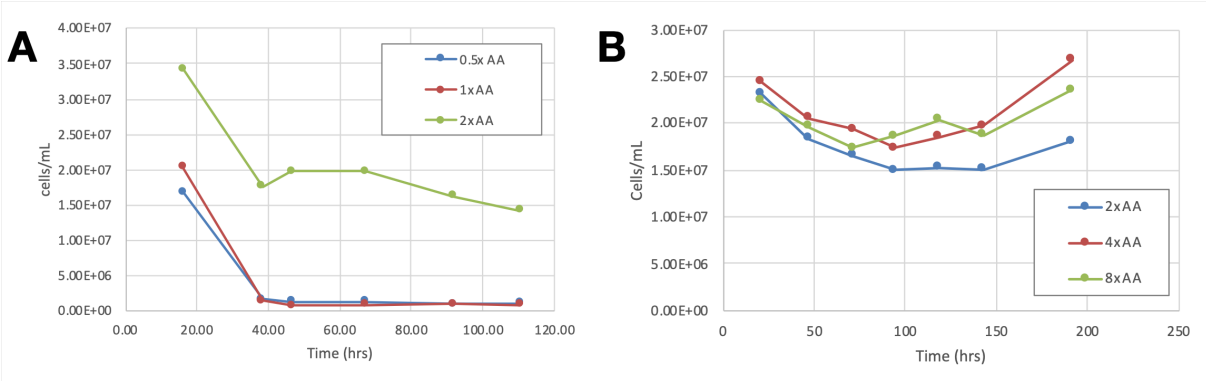


Figure 4.1: Optimization of supplemental additives needed for auxotrophic growth in phosphate-limited chemostats. (A) Initial test of 0.5X, 1X, and 2X concentrations of supplemental additives. (B) A 4X concentration of additives is large enough to prevent growth limitation. 1X = 20mg/L uracil, 20mg/L L-histidine HCl, 20mg/L L-tryptophan, and 1g L-leucine.

continuous culture, both the chemostat supplemented with 0.5X and 1X additives had effectively washed out (Figure 4.1A). The chemostat supplemented with a 2X concentration of additives was able to support growth of the haploid, but this was not sufficient to determine that this concentration was not limiting.

I again grew the haploid strains in three concentrations of additives, this time 2X, 4X, and 8X the standard recipe amounts. In this experiment, all three chemostats were able to support growth of the haploid strain, however the density of the 2X concentration culture was more dilute than the 4X and 8X cultures (Figure 4.1B). This revealed that the 2X concentration was still limiting growth more than the amount of phosphate in the media. Both the 4X and 8X cultures grew at approximately the same density, indicating both concentrations put the additives in enough excess to prevent growth limitation. While we could have further tested the concentrations required of individual additives to see which particular nutrient was limiting growth, we proceeded with optimizing dilution rate required for growth of the aneuploid strains in phosphate-limitation.

4.2.2 Optimization of chemostat-culturing of aneuploid strains

Growth into stationary phase has been shown to induce the environmental stress response in yeast (Gasch et al. 2000). We chose to grow the aneuploid and euploid strains in a phosphate-limited chemostat to control the growth rate of all strains at once. Most of the aneuploid strains showed some type of growth defect, and some of the strains were extremely slow growing, taking more than a day to reach saturation in batch culture. For this experiment, we had to carefully titrate the dilution rate of the cultures to be high enough to keep faster growing euploid strains in log-phase growth, but low enough to allow for the growth of the slowest growing strains.

Knowing the standard dilution rate of 0.17 was too high for most of the aneuploid strains, I first tried a dilution rate of 0.14, slightly below what was used for the tests of supplement concentrations. Strains A6, A7, and A18 did not reach saturation in the 36 hours of batch growth prior to turning on chemostat pumps, though strains A21, A20, and A2 were able to maintain growth at this dilution rate. A subsequent attempt with the same dilution rate resulted in all but one strain washing out. After these attempts, I found that a dilution rate of 0.11 allowed all of the aneuploid strains to maintain growth in the chemostat while also keeping the OD of all strains approximately between 0.5 and 1.0. Additionally, chemostats with strains with the highest degree of aneuploidy, A6 and A7, were inoculated 60-72 hours prior to turning the pumps on, while the remaining chemostats were inoculated 36-48 hours prior (Supplementary Table 4.1).

4.2.3 ESR signature is lost in aneuploids grown at a constant growth rate

Once strains were growing at steady state in the chemostat, defined as having maintained a constant cell density, cultures were harvested and sent to the Amon Lab at MIT for DNA, RNA, and ribosome extractions (see Materials and Methods). The ESR is a gene-expression

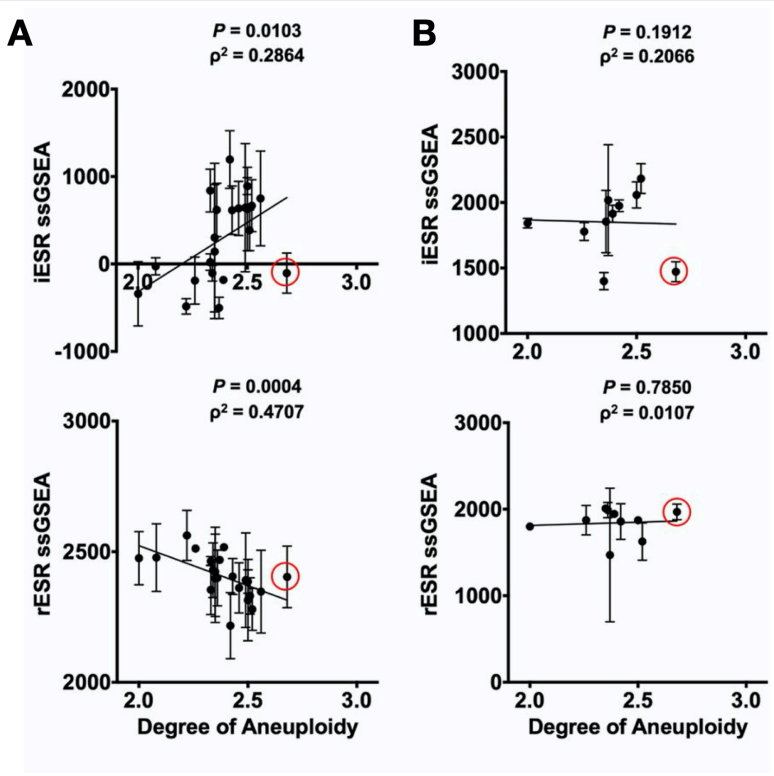


Figure 4.2: Aneuploid strains lose ESR signature when growth rate is controlled. (A) Correlations between iESR ssGSEA projections and degree of aneuploidy (Spearman, ρ = 0.2864, P = 0.0103) and rESR ssGSEA projections and degree of aneuploidy (Spearman, ρ = 0.4707, P = 0.0004) in batch culture grown strains. (B) Correlations between iESR ssGSEA projections and degree of aneuploidy (Spearman, ρ = 0.1912, P = 0.2066) and rESR ssGSEA projections and degree of aneuploidy (Spearman, ρ = 0.0107, P = 0.7850) for aneuploid strains grown to steady state in phosphate-limited chemostats. Error bars represent SD from the mean. The aneuploid strain circled in red does not display ESR. Figure panels were generated by Allegra Terhorst in the Amon Lab and are adapted from Terhorst et al. 2020.

pattern defined by a set of up- and downregulated genes. The Amon lab used RNA extracted from the chemostat samples to assess the degree to which the ESR was activated in the aneuploid cells compared to the euploid controls. In batch culture growth, a strong correlation exists between the degree of aneuploidy, calculated as the fraction of base pairs in the aneuploid strain relative to a haploid control strain, and the ESR (Figure 4.2A). When controlled for growth rate in the chemostat, this correlation disappears (Figure 4.2B). Because the ESR is defined by

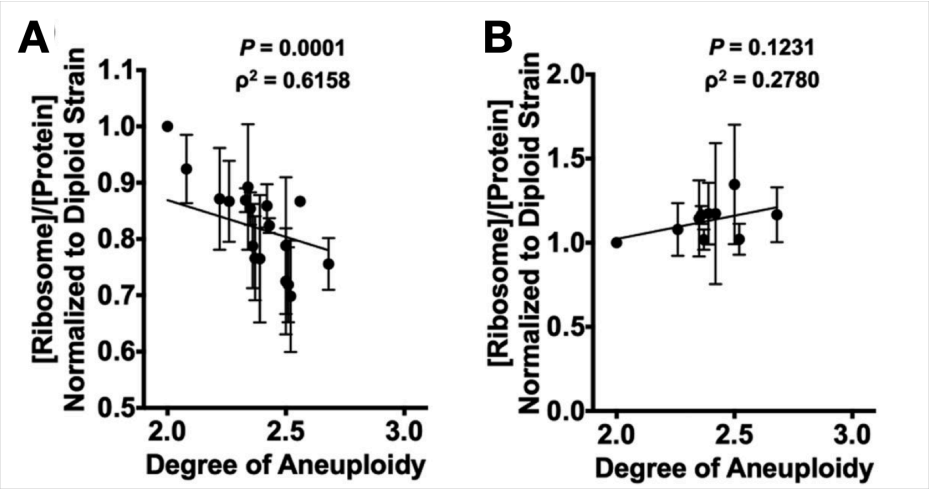


Figure 4.3: Ribosome loss in aneuploid cells is a result of ESR induction. (A) Correlation between [ribosome]/[protein] and degree of aneuploidy (ρ = 0.6158, P = 0.0001, Spearman) in batch culture grown strains. The calculated values were normalized to the [ribosome]/[protein] of a diploid control. (B) Correlation between [ribosome]/ [protein] and degree of aneuploidy (ρ = 0.2780, P = 0.1231, Spearman) for aneuploid strains grown to steady state in phosphate-limited chemostats. The calculated values were normalized to the [ribosome]/[protein] of a diploid control. Error bars represent SD from the mean of experimental replicates. Figure panels were generated by Allegra Terhorst in the Amon Lab and are adapted from Terhorst et al. 2020.

differential expression between aneuploid cells and the euploid controls, this suggests that the ESR is largely dependent on the proliferation rate, rather than the genomic content, of aneuploid cells.

4.2.4 ESR induces ribosome loss in aneuploid cells

Aneuploid cells have been shown to display a decreased cell density relative to euploid controls (Tsai et al. 2019). In this study, this observation was attributed to the protein imbalances prevalent in aneuploid cells causing a hypo-osmotic-like stress that led to swelling of the cell and subsequent cell density changes. The particular genes downregulated during the ESR hint that the decrease in cell density may be due to another cause. Ribosome fraction of the proteome is a large contributor to density of the cytoplasm (Delarue et al. 2018). Genes downregulated as a

part of the ESR include both ribosomal proteins and genes involved in ribosome biogenesis. If the decrease in cellular density was due to a loss of ribosomes, aneuploid cells would display fewer ribosomes than euploid controls. Furthermore, as the ESR was largely eliminated when aneuploid and euploid cells were matched in growth rate, if ribosome loss was due to the ESR, the depletion of ribosomes in aneuploid strains would also disappear.

Ribosomes were isolated from batch culture grown aneuploid and euploid cells at the Amon Lab using a modified protocol to enrich for assembled ribosomes (see Materials and Methods). Comparison of aneuploid and euploid strains showed that ribosomes did indeed make up a smaller fraction of the proteome in the aneuploid strains versus euploid controls (Figure 4.3A). The depletion of ribosomes in aneuploid strains was also correlated with the degree of aneuploidy. Cells harvested from growth rate controlled chemostats were also assayed for ribosome content. When the proliferation rate of euploid and aneuploid cells was equalized, the correlation of ribosome depletion with degree of aneuploidy was eliminated (Figure 4.3B). This suggests that the decrease in cellular density is a result of a loss of ribosomes which in turn is caused by induction of the ESR.

4.3 Discussion

The unbalanced nature of an aneuploid cell's genome results in a host of problems for the organism. Previous studies have attempted to characterize a generalized phenotype observed in aneuploid cells that may result in the proliferative delays seen in these cells. This response, known as the environmental stress response, is a generic transcriptional change observed in most aneuploid cells. During the ESR, ~900 genes are differentially expressed compared to euploid controls in the same proliferative state, including the upregulation of chaperones, amino acid transporters, and proteins with roles in endocytosis and proteasome activity, as well as the

downregulation of transcription and translation machinery, including ribosomal proteins. This response has been seen not only in yeast cells, but also aneuploid mammalian cells, indicating a conserved response across species (Sheltzer et al. 2012). Furthermore, the ESR has been shown to be a result of slowed cell division, which in turn is highly correlated to the degree of aneuploidy in a cell's genome.

Cell proliferation is almost always affected in stressful growth conditions, highlighting the generalized nature of the ESR signature, though the genes down-regulated during the ESR tend to have a tighter correlation with proliferation than the up-regulated ESR genes. Because down-regulated ESR genes are enriched for those involved in translation and transcription, it makes sense that these processes would need to be tightly controlled when growth is slowed, as failure to do so would result in heightened cell size increase prior to division. The down-regulation of genes in the ESR may be more important for preventing downstream effects of growth continuing while proliferation has slowed or stopped, while the up-regulated genes may be more tuned to the particular stress experienced by the cell. For example, upregulation of chaperone proteins may be critical to survival in high temperature to assist with a heightened need for protein folding and stabilization pathways, while amino acid transporters may be upregulated more in environments depleted of specific essential nutrients.

This work represents an important part of the larger question of how aneuploidy results in cellular fitness defects. Previously, while the ESR was known to correlate with the severity of aneuploidy present in a cell, it was relatively unknown exactly why aneuploidy resulted in a mounted stress response. In this study, we were able to control for a major characteristic of aneuploid cells— the decreased proliferation that was also correlated with the degree of aneuploidy. By growing complex aneuploid strains in a phosphate-limited chemostat, we could artificially reduce the growth rate of euploid controls to the same level as all aneuploid cells.

This completely eliminated the correlation of ESR with the degree of aneuploidy, indicating that the ESR is largely a result of slowed proliferation in aneuploid cells and other cells experiencing environmental stress.

We were also able to shed light on the observation that aneuploid cells showed a dramatic decrease in cytoplasmic density (Tsai et al. 2019). This was thought to be related to the activation of hypoosmotic stress responses, though the effect sizes of these two responses seemed inconsistent with this link. Because ribosomes make up a huge fraction of cytoplasmic proteins, and the ESR involves downregulation of both ribosomal proteins and ribosome biogenesis genes, we tested whether the decrease in cytoplasmic density was actually due to the ESR. Normalizing cell proliferation in the chemostat had equalized the effect of the ESR between euploid and aneuploid cells. We assayed the fraction of total cellular protein that ribosomes account for in chemostat-grown and batch culture-grown populations of aneuploid and euploid cells. In batch culture, as the degree of aneuploidy increased in a cell, and thus the strength of the ESR increased, cells exhibited a lower fraction of ribosomes in the total cellular protein concentration. Controlling the ESR through growth in the chemostat eliminated this trend, indicating that aneuploidy itself did not result in a decreased number of ribosomes, but ribosomes are actually lost as a consequence of mounting the ESR.

This work also provided a useful optimization of growing difficult strains in the chemostats. We have rarely used auxotrophic strains in the chemostat, and in cases of single auxotrophies, it is tempting to use previously published concentrations of additives for batch culture. In this study, we highlight the importance of verifying the nutrient limiting growth in a chosen condition. Concentrations of leucine, histidine, uracil, and tryptophan published for use in batch culture were clearly insufficient for even the haploid control to grow at our typical chemostat growth rate. Proceeding when these nutrients were not present in excess would result

in additional stressors placed on cells that would certainly confound all downstream analyses. With this work as a baseline, we could experimentally determine the necessary concentrations of a variety of other common auxotrophies, though whether the presence of multiple auxotrophies alters required concentrations remains to be seen.

4.4 Materials and Methods

Strains used in this study: Strains used in this study are listed in Supplementary Table 4.1. All strains were provided by Arzu Sandikci from the Amon Lab. Complex aneuploid strains were constructed in the Li Lab while at the Stowers Institute for Medical Research and are described in (Pavelka et al. 2010).

Chemostat growth media: Standard nutrient-limited chemostat media recipes are provided at http://dunham.gs.washington.edu/protocols.shtml. Standard concentrations of histidine, uracil, tryptophan, and leucine were obtained from the Methods in Yeast Genetics and Genomics Cold Spring Harbor Laboratory Manual.

Chemostat cultures and sampling for RNA and DNA: Chemostats were filled with phosphate-limited media and inoculated with 2mL of overnight culture. Strains A6 and A7 were grown in batch for three days and all other strains for two days prior to turning on peristaltic pumps. Pumps were set at a dilution rate of ~0.11 culture volumes/hour, unless otherwise indicated. Samples were taken for dilution rate calculation and density measurements every day, with density measurements taken on a spectrophotometer and counted manually with a hemocytometer. Cultures were determined to have reached steady state when density measurements differed from the previous sampling by less than 5%.

When cultures had reached steady state, samples were taken for DNA, RNA, and ribosome preparations. DNA samples were collected by pelleting 5mL culture and freezing

at -20°C for storage until harvesting. DNA samples were sequenced on an Illumina HiSeq2000. RNA samples were collected by vacuum filtering 50mL culture onto 47mm nylon filters (Fisher Scientific, cat. no. R04SP04700). The filters were rolled using sterile tweezers, placed into a 15mL Falcon tube, and snap frozen in liquid nitrogen. Total RNA was then sequenced using Illumina Truseq and then by Nextera or Roche KAPA.

Single-Sample Gene Set Enrichment Analysis (ssGSEA): Reads were aligned to the *S. cerevisiae* transcriptome using STAR (version 2.5.3a, Dobin et al. 2013). Gene expression was quantified using RSEM (version 1.3.0, Li and Dewey 2011). Values were $\log_2$ transformed and used to prepare gene cluster text files for ssGSEA (version 7.7, Subramanian et al 2005, Barbie et al. 2009). Projections for ESR are as described in (Gasch et al. 2012). Sort order had a minor effect on ssGSEA. ssGSEA analysis was performed in the Amon Lab at MIT by Allegra Terhorst.

Ribosomal content analysis: Samples for ribosome processing were gathered by spinning down 50mL culture and snap freezing in liquid nitrogen. Upon arrival at MIT, samples were resuspended in lysis buffer (20mM Hepes pH 7.4, 100mM potassium acetate, 2mM magnesium acetate, 3mM dithiothreitol (DTT), ethylenediaminetetraacetic acid-free protease inhibitor [Roche, cat. No. 11836170001]) and treated with 0.5mg/mL zymolase. Cells were lysed twice with a French press and spun at 19,000 rpm for 20 minutes at 4°C to remove cellular debris. Protein concentration of lysate was determined by Bradford assay. 80S ribosomes were purified by overlaying 10mL lysate onto 15mL prechilled 4°C sucrose solution (30% sucrose, 20mM Hepes pH 7.4, 500mM potassium acetate, 2mM magnesium acetate, 3mM DTT) and spinning at 50,000 rpm for 4 hours at 4°C. Pellets were dried upside down for 10 min and resuspended in 1mL lysis buffer. Concentration of purified assembled ribosomes was determined by 260 nm

absorbance on a Nanodrop. Cell number counts for each culture were determined using a Beckman Multisizer 3 Coulter Counter to determine ribosome fraction per cell.

Chapter 5. Conclusions and future directions

Throughout this thesis, I have presented my work investigating the consequences aneuploidy has on a cellular and organismal level. In Chapter 2, I discussed the application and optimization of the Telomeric Amplicon (Tamp) collection to conditions that are normally detrimental to the growth of aneuploid cells. In Chapter 3, I attempted to narrow in on two particular breakpoints observed from previous Tamp competitions and further developed a method for assessing the contribution of individual genes to fitness effects of large amplifications. Finally, in Chapter 4, I explored the more generalized phenotypes of aneuploid cells and traced the stress response exhibited by aneuploid cells back to the proliferation defects shared by most aneuploid cells. In this final chapter, I will discuss my work in the context of the field at large and explore possible future research that may follow from this line of study.

5.1 The use of the Tamp collection in additional environments

The Tamp collection remains a powerful tool to assess how each specific region of the genome may contribute to fitness effects observed in cells bearing large amplifications. In this work I have adapted the use of the collection for additional experimental conditions and improved the analysis pipeline to be more flexible for future work. The original construction of the Tamp pool was centered around its intended application for assessing fitness effects of aneuploidy in phosphate-, glucose-, and sulfate-limited chemostats (Sunshine et al. 2015). This built on previous work in the lab assessing the fitness effects of changing the copy number of single genes across the genome (Payen et al. 2016). Adapting this approach for use in a vastly different condition required significant changes to the analysis pipeline published previously.

The initial experiments in nutrient limitation were based on the observation that yeast grown in this condition often adapt through the amplification of large segments of a

chromosome. In some cases, recurrent amplifications and variable boundaries of these amplifications, coupled with single gene deletion and amplification data, strongly suggested that fitness effects may be driven by single gene copy number changes. And use of the Tamp collection largely showed this to be true— the landscape of fitness effects seen across the three nutrient limitations was mostly flat, punctuated by step points hypothesized to contain genes that disproportionately contribute to fitness when amplified. The method chosen to call breakpoints in this data set was determined by this observation. DNAcopy, an R program used to call copy number changes in DNA copy number data, fits a series of horizontal lines denoting the copy-number, or in this case fitness, of each segment of DNA.

My work with the Tamp collection focused on conditions that aneuploid strains tolerated poorly. Because aneuploidy had been associated with generalized stress responses, discussed more in Chapter 4, I sought to determine whether fitness consequences of aneuploidy were also determined by the specific genes amplified rather than simply the state of being aneuploid. For this case, a model of calling fitness step points that by default fit a flat line broken up by steps may not fully explain more complex effects on fitness as amplification size gradually increased. To explore possibilities of modeling our fitness data, we worked with Lucy Gao and Daniela Witten of the UW Biostatistics department. We devised an approach to fit both a piecewise constant model, like in the DNAcopy program, and a simple linear regression and tested which model resulted in a better fit of our fitness data. Ultimately, we found that differences between our experimental protocols, specifically batch culture versus chemostat growth, and the level of noise present in each dataset, limited the usefulness of this approach. We found that in some cases where a linear model provided the best fit, step points obvious to the eye that resulted in high magnitude changes of fitness were masked in our final candidate list. Conversely,

chromosome arms with high Tamp coverage were best fit by the piecewise constant model, yet all called breakpoints were of a very low magnitude and seemed a likely artifact of overfitting. Due to these challenges, we abandoned this approach to identifying breakpoints, though we still employed the piecewise constant model and filtered step points for magnitude to prevent overfitting. In the future, however, I believe the method for calling potential breakpoints could be revisited. The piecewise constant model has a similar flaw as the DNAcopy approach in that breakpoints are called on the background of flat steps. This sets up a false dichotomy where fitness must either be determined by single genes or by the length of the amplification. Based on what I have observed of the consequences of large amplifications, the likely reason is that both explanations are true to some extent. In this case, fitness may decrease as the amplification length increases, and on this background of a linear relationship, single genes may drive condition-specific fitness effects of particular amplifications.

5.2 Further optimization of our driver gene validation approach

To what extent individual genes contribute to the fitness effects of large amplifications is still a largely open question. In Chapter 3, I attempted to determine the effects of single genes on the fitness of a large amplification of chromosome 10 in phosphate- and glucose-limited chemostats. The selection of these breakpoints was in part convenience— the neighboring regions could be assayed using the same large amplification— but also due to the lack of clear candidate genes in the region. With this work, I hoped to distinguish among several possible explanations. First, I considered whether a single gene was responsible for fitness effects of the amplification. Data collected on the fitness effects of single-gene deletions and amplifications contained several missing datapoints that could be the result of highly deleterious mutations, making them ideal candidates in this study. Second, the fitness effects of each gene across the

breakpoint could have an additive effect. In this case, had the coverage of the Tamp collection contained amplifications starting at every possible gene across the breakpoint, we would see a gradual decrease in fitness that was merely masked by a lapse in coverage.

Lastly, the most interesting possible explanation is that specific gene amplifications result in a negative effect only when present in the context of an aneuploid cell. We know that aneuploidy results in both condition- and chromosome-specific consequences as well as a generalized stress phenotype. It is possible that certain gene amplifications are consequential only in the presence of this stress response. Based on work in Chapter 4 investigating when and why aneuploid cells mount the ESR, we may be able to compare the effects of single gene copy number changes in conditions that also activate the ESR to begin to look at these possible interactions.

5.3 Development of a genome-wide truncation collection

The work presented in this thesis is heavily focused on the effects that large amplifications have on cellular fitness. One future approach to investigating the effects of aneuploidy is the development of a parallel collection of systematic chromosome truncations. Evolved populations of yeast have displayed large deletions in addition to the amplifications discussed in this work. For example, in glucose limitation, multiple evolved clones displayed a genomic rearrangement consisting of the amplification of part of chromosome 4 and the deletion of one copy of a region of chromosome 15 (Dunham et al. 2002). This translocation event results in an increase in copy number of the hexose transporter *HXT6* which likely drives the amplification of chromosome 15, though it is unclear whether the deletion on chromosome 15 also provides an adaptive benefit or merely serves as a convenient and low-cost site for the accompanying amplification. These types of chromosome rearrangements are incredibly

common in cancer cell lines, and further characterization of the consequences and benefits of chromosome deletions may prove enlightening in the cancer field.

Introduction of a chromosome fragmentation vector with slight alterations in component orientation and the omission of a centromere on the plasmid can result in a recombination event that truncates the native chromosome arm rather than amplifying off of it (Morrow et al. 1997). Using the same approach taken with the Tamp collection, a telomeric deletion, or Tdel, collection could be made off the background of the yeast knockout collection where every strain contained a barcoded truncation of a chromosome tiled across the genome. However, deletion events generally have a larger effect size that amplifications, even in a diploid yeast. Copy number changes from two to three create an inherently smaller effect than from two to one. As all truncations would extend to the telomere, any haploinsufficient gene on a chromosome arm would preclude formation of truncation strains upstream of their location. Additionally, while selection of the Tamp could be maintained by a selectable marker on the amplification, it is harder to select for the absence of part of a chromosome in yeast. If the truncation is deleterious enough, an amplification of the entire native chromosome could confer enough benefit to overcome the effects. The extent to which this could happen would likely depend only on the balance between deleterious effects of a whole chromosome or a partial truncation. A pooled approach to truncations would need to address this possibility, perhaps through a dosage-dependent marker or introduction of a fluorescent chromosome marker that can be selected for appropriate copy number.

5.4 Contribution of genomic context to the fitness of aneuploid strains

The specific types of aneuploidy that we have focused on here are single large amplifications and deletions. Yet aneuploid events in nature are often complex and variable.

Tumors and cancer cell lines display massive structural and copy number alterations that together result in highly proliferative populations. By investigating the effects of individual amplifications on cellular fitness in different conditions, we can begin to understand how karyotype influences proliferative success, though how each amplification behaves in combination with other rearrangements becomes incredibly complex very fast.

The Tamp pool results also may have been impacted by interactions of single amplifications and background genetic factors. The pool is derived from the yeast knockout collection, in which strains have been shown to harbor additional mutations and genomic rearrangements to compensate for the loss of even non-essential genes. Our breakpoint calling pipeline relies on the patterns of fitness exhibited by many strains across chromosome arms, and as a result we can limit the impact that strain-specific effects have on our analysis. However, in our validation assay, where fitness of each strain was compared to a much smaller pool, differences in individual deletion strain karyotypes could have been a fatal source of noise in the experiment.

Genomic context may have contributed to Tamp fitness in another way. Throughout our analysis of the Tamp competition, we focused on the effects that individual genes had on the overall fitness conferred by the amplification. But the genome is made up of more than just genes, and we left the contributions of other genomic factors largely unexplored. Further analysis of the results of these and future studies should take into account how non-genic elements may impact fitness. For example, proliferation rate by necessity depends to some extent on how fast the genome can be replicated. Our pool relies on native origins of replication to populate the synthetic amplification. In regions of the genome naturally low in early firing origins, the Tamp may be slow to replicate as it lacks the benefit of ARSs proceeding from the opposite direction on regions not present on the amplification. In an area where we might suspect this occurrence,

we could compare the fitness of individual Tamps to matched strains where the Tamp is supplemented by an additional origin sequence.

Bibliography

Ahn K, Gotay N, Andersen TM, Anvari AA, Gochman P, Lee Y, Sanders S, Guha S, Darvasi A, Glessner JT, et al. 2014. High rate of disease-related copy number variations in childhood onset schizophrenia. Mol Psychiatry. 19(5):568–572. doi:10.1038/mp.2013.59.

Altenburg T, Goldenbogen B, Uhlendorf J, Klipp E. 2019. Osmolyte homeostasis controls single-cell growth rate and maximum cell size of *Saccharomyces cerevisiae*. npj Syst Biol Appl. 5(1):34. doi:10.1038/s41540-019-0111-6.

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, et al. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 462(7269):108–112. doi:10.1038/nature08460.

Bassett AS, Lowther C, Merico D, Costain G, Chow EWC, van Amelsvoort T, McDonald-McGinn D, Gur RE, Swillen A, Van den Bree M, et al. 2017. Rare Genome-Wide Copy Number Variation and Expression of Schizophrenia in 22q11.2 Deletion Syndrome. Am J Psychiatry. 174(11):1054–1063. doi:10.1176/appi.ajp.2017.16121417.

Bastien P, Blaineau C, Pagès M. 1992. Molecular karyotype analysis in *Leishmania*. Subcell Biochem. 18:131–187. doi:10.1007/978-1-4899-1651-8_5.

Benhra N, Barrio L, Muzzopappa M, Milán M. 2018. Chromosomal Instability Induces Cellular Invasion in Epithelial Tissues. Dev Cell. 47(2):161-174.e4. doi:10.1016/j.devcel.2018.08.021.

Bonassi S, Hagmar L, Strömberg U, Montagud AH, Tinnerberg H, Forni A, Heikkilä P, Wanders S, Wilhardt P, Hansteen IL, et al. 2000. Chromosomal aberrations in lymphocytes predict human cancer independently of exposure to carcinogens. European Study Group on Cytogenetic Biomarkers and Health. Cancer Res. 60(6):1619–1625.

Boveri T. 2008. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. J Cell Sci. 121 Suppl 1:1–84. doi:10.1242/jcs.025742.

Braun R, Ronquist S, Wangsa Darawalee, Chen H, Anthuber L, Gemoll T, Wangsa Danny, Koparde V, Hunn C, Habermann JK, et al. 2019. Single Chromosome Aneuploidy Induces Genome-Wide Perturbation of Nuclear Organization and Gene Expression. Neoplasia. 21(4):401–412. doi:10.1016/j.neo.2019.02.003.

Brewer BJ, Payen C, Raghuraman MK, Dunham MJ. 2011. Origin-dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons. PLoS Genet. 7(3):e1002016. doi:10.1371/journal.pgen.1002016.

Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet. 38(9):1043–1048. doi:10.1038/ng1861.

Chailangkarn T, Noree C, Muotri AR. 2018. The contribution of GTF2I haploinsufficiency to Williams syndrome. Mol Cell Probes. 40:45–51. doi:10.1016/j.mcp.2017.12.005.

Chan JE, Kolodner RD. 2011. A genetic and structural study of genome rearrangements mediated by high copy repeat Ty1 elements. PLoS Genet. 7(5):e1002089. doi:10.1371/journal.pgen.1002089.

Chance PF, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, Lupski JR. 1994. Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. Hum Mol Genet. 3(2):223–228. doi:10.1093/hmg/3.2.223.

Chunduri NK, Storchová Z. 2019. The diverse consequences of aneuploidy. Nat Cell Biol. 21(1):54–62. doi:10.1038/s41556-018-0243-8.

Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, Nezi L, Protopopov A, Chowdhury D, Pellman D. 2012. DNA breaks and chromosome pulverization from errors in mitosis. Nature. 482(7383):53–58. doi:10.1038/nature10802.

Dai C, Sampson SB. 2016. HSF1: Guardian of Proteostasis in Cancer. Trends Cell Biol. 26(1):17–28. doi:10.1016/j.tcb.2015.10.011.

Delarue M, Brittingham GP, Pfeffer S, Surovtsev IV, Pinglay S, Kennedy KJ, Schaffer M, Gutierrez JI, Sang D, Poterewicz G, et al. 2018. mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding. Cell. 174(2):338-349.e20. doi:10.1016/j.cell.2018.05.042.

Dephoure N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, Torres EM. 2014. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. Elife. 3:e03023. doi:10.7554/eLife.03023.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15–21. doi:10.1093/bioinformatics/bts635.

Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 99(25):16144–16149. doi:10.1073/pnas.242624799.

Ecker RE, Schaechter M. 1963. Ribosome content and the rate of growth of *Salmonella typhimurium*. Biochimica et Biophysica Acta (BBA) - Specialized Section on Nucleic Acids and Related Subjects. 76:275–279. doi:10.1016/0926-6550(63)90040-9.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 11(12):4241–4257. doi:10.1091/mbc.11.12.4241.

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 4(12):e1000303. doi:10.1371/journal.pgen.1000303.

Gresham D, Usaite R, Germann SM, Lisby M, Botstein D, Regenberg B. 2010. Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the *GAP1* locus. Proc Natl Acad Sci U S A. 107(43):18551–18556. doi:10.1073/pnas.1014023107.

Haarer B, Aggeli D, Viggiano S, Burke DJ, Amberg DC. 2011. Novel interactions between actin and the proteasome revealed by complex haploinsufficiency. PLoS Genet. 7(9):e1002288. doi:10.1371/journal.pgen.1002288.

Hardy PA, Zacharias H. 2005. Reappraisal of the Hansemann-Boveri hypothesis on the origin of tumors. Cell Biol Int. 29(12):983–992. doi:10.1016/j.cellbi.2005.10.001.

Hart SFM, Pineda JMB, Chen C-C, Green R, Shou W. 2019. Disentangling strictly self-serving mutations from win-win mutations in a mutualistic microbial community. Elife. 8. doi:10.7554/eLife.44812.

Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet. 2(4):280–291. doi:10.1038/35066065.

Helbig I, Swinkels MEM, Aten E, Caliebe A, van 't Slot R, Boor R, von Spiczak S, Muhle H, Jähn JA, van Binsbergen E, et al. 2014. Structural genomic variation in childhood epilepsies with complex phenotypes. Eur J Hum Genet. 22(7):896–901. doi:10.1038/ejhg.2013.262.

Hoffman CS, Winston F. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene. 57(2–3):267–272. doi:10.1016/0378-1119(87)90131-4.

Inoue K, Dewar K, Katsanis N, Reiter LT, Lander ES, Devon KL, Wyman DW, Lupski JR, Birren B. 2001. The 1.4-Mb *CMT1A* duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. Genome Res. 11(6):1018–1033. doi:10.1101/gr.180401.

Kafri M, Metzl-Raz E, Jona G, Barkai N. 2016. The Cost of Protein Production. Cell Rep. 14(1):22–31. doi:10.1016/j.celrep.2015.12.015.

Kim T-M, Xi R, Luquette LJ, Park RW, Johnson MD, Park PJ. 2013. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. Genome Res. 23(2):217–227. doi:10.1101/gr.140301.112.

Kirov G, Rees E, Walters JTR, Escott-Price V, Georgieva L, Richards AL, Chambert KD, Davies G, Legge SE, Moran JL, et al. 2014. The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry. 75(5):378–385. doi:10.1016/j.biopsych.2013.07.022.

Koszul R, Caburet S, Dujon B, Fischer G. 2004. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. EMBO J. 23(1):234–243. doi:10.1038/sj.emboj.7600024.

Koszul R, Dujon B, Fischer G. 2006. Stability of large segmental duplications in the yeast genome. Genetics. 172(4):2211–2222. doi:10.1534/genetics.105.048058.

Koufaris C, Sismani C. 2015. Modulation of the genome and epigenome of individuals susceptible to autism by environmental risk factors. Int J Mol Sci. 16(4):8699–8718. doi:10.3390/ijms16048699.

Kurotaki N, Shen JJ, Touyama M, Kondoh T, Visser R, Ozaki T, Nishimoto J, Shiihara T, Uetake K, Makita Y, et al. 2005. Phenotypic consequences of genetic variation at hemizygous alleles: Sotos syndrome is a contiguous gene syndrome incorporating

coagulation factor twelve (FXII) deficiency. Genet Med. 7(7):479–483. doi:10.1097/01.gim.0000177419.43309.37.

Kvitek DJ, Sherlock G. 2011. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet. 7(4):e1002056. doi:10.1371/journal.pgen.1002056.

Laffitte M-CN, Leprohon P, Papadopoulou B, Ouellette M. 2016. Plasticity of the *Leishmania* genome leading to gene copy number variations and drug resistance. F1000Res. 5:2350. doi:10.12688/f1000research.9218.1.

Lara-Gonzalez P, Westhorpe FG, Taylor SS. 2012. The spindle assembly checkpoint. Curr Biol. 22(22):R966-980. doi:10.1016/j.cub.2012.10.006.

Lebreton A, Saveanu C, Decourty L, Rain J-C, Jacquier A, Fromont-Racine M. 2006. A functional network involved in the recycling of nucleocytoplasmic pre-60S factors. J Cell Biol. 173(3):349–360. doi:10.1083/jcb.200510080.

Lemoine FJ, Degtyareva NP, Lobachev K, Petes TD. 2005. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. Cell. 120(5):587–598. doi:10.1016/j.cell.2004.12.039.

Lengauer C, Kinzler KW, Vogelstein B. 1998. Genetic instabilities in human cancers. Nature. 396(6712):643–649. doi:10.1038/25292.

Leprohon P, Légaré D, Raymond F, Madore E, Hardiman G, Corbeil J, Ouellette M. 2009. Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. Nucleic Acids Res. 37(5):1387–1399. doi:10.1093/nar/gkn1069.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323. doi:10.1186/1471-2105-12-323.

Linder RA, Greco JP, Seidl F, Matsui T, Ehrenreich IM. 2017. The Stress-Inducible Peroxidase *TSA2* Underlies a Conditionally Beneficial Chromosomal Duplication in *Saccharomyces cerevisiae*. G3 (Bethesda). 7(9):3177–3184. doi:10.1534/g3.117.300069.

Lindsley DL, Sandler L, Baker BS, Carpenter AT, Denell RE, Hall JC, Jacobs PA, Miklos GL, Davis BK, Gethmann RC, et al. 1972. Segmental aneuploidy and the genetic gross structure of the *Drosophila* genome. Genetics. 71(1):157–184.

Lupski JR, Wise CA, Kuwano A, Pentao L, Parke JT, Glaze DG, Ledbetter DH, Greenberg F, Patel PI. 1992. Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. Nat Genet. 1(1):29–33. doi:10.1038/ng0492-29.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, et al. 2008. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet. 82(2):477–488. doi:10.1016/j.ajhg.2007.12.009.

Mefford HC, Eichler EE. 2009. Duplication hotspots, rare genomic disorders, and common disease. Curr Opin Genet Dev. 19(3):196–204. doi:10.1016/j.gde.2009.04.003.

Miller AW, Befort C, Kerr EO, Dunham MJ. 2013. Design and use of multiplexed chemostat arrays. J Vis Exp.(72):e50262. doi:10.3791/50262.

Mohajeri K, Cantsilieris S, Huddleston J, Nelson BJ, Coe BP, Campbell CD, Baker C, Harshman L, Munson KM, Kronenberg ZN, et al. 2016. Interchromosomal core duplicons drive both evolutionary instability and disease susceptibility of the Chromosome 8p23.1 region. Genome Res. 26(11):1453–1467. doi:10.1101/gr.211284.116.

Morrow DM, Connelly C, Hieter P. 1997. “Break copy” duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. Genetics. 147(2):371–382.

Nagaoka SI, Hassold TJ, Hunt PA. 2012. Human aneuploidy: mechanisms and new insights into an age-old problem. Nat Rev Genet. 13(7):493–504. doi:10.1038/nrg3245.

Niwa O, Tange Y, Kurabayashi A. 2006. Growth arrest and chromosome instability in aneuploid yeast. Yeast. 23(13):937–950. doi:10.1002/yea.1411.

Oltmann J, Heselmeyer-Haddad K, Hernandez LS, Meyer R, Torres I, Hu Y, Doberstein N, Killian JK, Petersen D, Zhu YJ, et al. 2018. Aneuploidy, *TP53* mutation, and amplification of *MYC* correlate with increased intratumor heterogeneity and poor prognosis of breast cancer patients. Genes Chromosomes Cancer. 57(4):165–175. doi:10.1002/gcc.22515.

Pavan WJ, Hieter P, Sears D, Burkhoff A, Reeves RH. 1991. High-efficiency yeast artificial chromosome fragmentation vectors. Gene. 106(1):125–127. doi:10.1016/0378-1119(91)90576-W.

Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R. 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 468(7321):321–325. doi:10.1038/nature09529.

Payen C, Di Rienzi SC, Ong GT, Pogachar JL, Sanchez JC, Sunshine AB, Raghuraman MK, Brewer BJ, Dunham MJ. 2014. The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. G3 (Bethesda). 4(3):399–409. doi:10.1534/g3.113.009365.

Payen C, Sunshine AB, Ong GT, Pogachar JL, Zhao W, Dunham MJ. 2016. High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations. PLoS Genet. 12(10):e1006339. doi:10.1371/journal.pgen.1006339.

Perepnikhatka V, Fischer FJ, Niimi M, Baker RA, Cannon RD, Wang YK, Sherman F, Rustchenko E. 1999. Specific chromosome alterations in fluconazole-resistant mutants of *Candida albicans*. J Bacteriol. 181(13):4041–4049. doi:10.1128/JB.181.13.4041-4049.1999.

Potapova T, Gorbsky GJ. 2017. The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis. Biology (Basel). 6(1). doi:10.3390/biology6010012.

Puddu F, Herzog M, Selivanova A, Wang S, Zhu J, Klein-Lavi S, Gordon M, Meirman R, Millan-Zambrano G, Ayestaran I, et al. 2019. Genome architecture and stability in the *Saccharomyces cerevisiae* knockout collection. Nature. 573(7774):416–420. doi:10.1038/s41586-019-1549-9.

Rice AM, McLysaght A. 2017. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. Nat Commun. 8:14366. doi:10.1038/ncomms14366.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol. 29(1):24–26. doi:10.1038/nbt.1754.

Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, et al. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. Genome Res. 21(12):2129–2142. doi:10.1101/gr.122945.111.

Rutkowski TP, Schroeder JP, Gafford GM, Warren ST, Weinshenker D, Caspary T, Mulle JG. 2017. Unraveling the genetic architecture of copy number variants associated with schizophrenia and other neuropsychiatric disorders. J Neurosci Res. 95(5):1144–1160. doi:10.1002/jnr.23970.

Schaechter M, Maaloe O, Kjeldgaard NO. 1958. Dependency on medium and temperature of cell size and chemical composition during balanced grown of *Salmonella typhimurium*. J Gen Microbiol. 19(3):592–606. doi:10.1099/00221287-19-3-592.

Sears DD, Hegemann JH, Hieter P. 1992. Meiotic recombination and segregation of human-derived artificial chromosomes in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 89(12):5296–5300. doi:10.1073/pnas.89.12.5296.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, et al. 2007. Strong association of de novo copy number mutations with autism. Science. 316(5823):445–449. doi:10.1126/science.1138659.

Segal DJ, McCoy EE. 1974. Studies on Down's syndrome in tissue culture. I. Growth rates and protein contents of fibroblast cultures. J Cell Physiol. 83(1):85–90. doi:10.1002/jcp.1040830112.

Selmecki A, Bergmann S, Berman J. 2005. Comparative genome hybridization reveals widespread aneuploidy in *Candida albicans* laboratory strains. Mol Microbiol. 55(5):1553–1565. doi:10.1111/j.1365-2958.2005.04492.x.

Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. Science. 313(5785):367–370. doi:10.1126/science.1128242.

Selmecki A, Gerami-Nejad M, Paulson C, Forche A, Berman J. 2008. An isochromosome confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1. Mol Microbiol. 68(3):624–641. doi:10.1111/j.1365-2958.2008.06176.x.

Selmecki AM, Dulmage K, Cowen LE, Anderson JB, Berman J. 2009. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. PLoS Genet. 5(10):e1000705. doi:10.1371/journal.pgen.1000705.

Sheltzer JM. 2013. A transcriptional and metabolic signature of primary aneuploidy is present in chromosomally unstable cancer cells and informs clinical prognosis. Cancer Res. 73(21):6401–6412. doi:10.1158/0008-5472.CAN-13-0749.

Sheltzer JM, Ko JH, Replogle JM, Habibe Burgos NC, Chung ES, Meehl CM, Sayles NM, Passerini V, Storchova Z, Amon A. 2017. Single-chromosome Gains Commonly Function as Tumor Suppressors. Cancer Cell. 31(2):240–255. doi:10.1016/j.ccell.2016.12.004.

Sheltzer JM, Torres EM, Dunham MJ, Amon A. 2012. Transcriptional consequences of aneuploidy. Proc Natl Acad Sci U S A. 109(31):12644–12649. doi:10.1073/pnas.1209227109.

Simonetti G, Bruno S, Padella A, Tenti E, Martinelli G. 2019. Aneuploidy: Cancer strength or vulnerability? Int J Cancer. 144(1):8–25. doi:10.1002/ijc.31718.

Smukowski Heil CS, Large CRL, Patterson K, Hickey AS-M, Yeh C-LC, Dunham MJ. 2019. Temperature preference can bias parental genome retention during hybrid evolution. PLoS Genet. 15(9):e1008383. doi:10.1371/journal.pgen.1008383.

Sotillo R, Hernando E, Díaz-Rodríguez E, Teruya-Feldstein J, Cordón-Cardo C, Lowe SW, Benezra R. 2007. Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. Cancer Cell. 11(1):9–23. doi:10.1016/j.ccr.2006.10.019.

Sterkers Y, Lachaud L, Bourgeois N, Crobu L, Bastien P, Pagès M. 2012. Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. Mol Microbiol. 86(1):15–23. doi:10.1111/j.1365-2958.2012.08185.x.

Stopsack KH, Whittaker CA, Gerke TA, Loda M, Kantoff PW, Mucci LA, Amon A. 2019. Aneuploidy drives lethal progression in prostate cancer. Proc Natl Acad Sci U S A. 116(23):11390–11395. doi:10.1073/pnas.1902645116.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 102(43):15545–15550. doi:10.1073/pnas.0506580102.

Sunshine AB, Payen C, Ong GT, Liachko I, Tan KM, Dunham MJ. 2015. The fitness consequences of aneuploidy are driven by condition-dependent gene effects. PLoS Biol. 13(5):e1002155. doi:10.1371/journal.pbio.1002155.

Tatton-Brown K, Cole TR, Rahman N. 1993. Sotos Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, Amemiya A, editors. GeneReviews®. Seattle (WA): University of Washington, Seattle. [accessed 2020]. http://www.ncbi.nlm.nih.gov/books/NBK1479/.

Taxis C, Keller P, Kavagiou Z, Jensen LJ, Colombelli J, Bork P, Stelzer EHK, Knop M. 2005. Spore number control and breeding in *Saccharomyces cerevisiae*: a key role for a self-organizing system. J Cell Biol. 171(4):627–640. doi:10.1083/jcb.200507168.

Terhorst A, Sandikci A, Keller A, Whittaker CA, Dunham MJ, Amon A. 2020. The environmental stress response causes ribosome loss in aneuploid yeast cells. Proc Natl Acad Sci U S A. 117(29):17031–17040. doi:10.1073/pnas.2005648117.

Theisen A, Shaffer LG. 2010. Disorders caused by chromosome abnormalities. Appl Clin Genet. 3:159–174. doi:10.2147/TACG.S8884.

Thompson SL, Compton DA. 2011. Chromosome missegregation in human cells arises through specific types of kinetochore-microtubule attachment errors. Proc Natl Acad Sci U S A. 108(44):17974–17978. doi:10.1073/pnas.1109720108.

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science. 294(5550):2364–2368. doi:10.1126/science.1065810.

Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, Dunham MJ, Amon A. 2010. Identification of aneuploidy-tolerating mutations. Cell. 143(1):71–83. doi:10.1016/j.cell.2010.08.038.

Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A. 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science. 317(5840):916–924. doi:10.1126/science.1142210.

Treco DA, Winston F. 2008. Growth and manipulation of yeast. Curr Protoc Mol Biol. Chapter 13:Unit 13.2. doi:10.1002/0471142727.mb1302s82.

Tsai H-J, Nelliat AR, Choudhury MI, Kucharavy A, Bradford WD, Cook ME, Kim J, Mair DB, Sun SX, Schatz MC, et al. 2019. Hypo-osmotic-like stress underlies general cellular defects of aneuploidy. Nature. 570(7759):117–121. doi:10.1038/s41586-019-1187-2.

Ubeda J-M, Légaré D, Raymond F, Ouameur AA, Boisvert S, Rigault P, Corbeil J, Tremblay MJ, Olivier M, Papadopoulou B, et al. 2008. Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. Genome Biol. 9(7):R115. doi:10.1186/gb-2008-9-7-r115.

Usher CL, McCarroll SA. 2015. Complex and multi-allelic copy number variation in human disease. Brief Funct Genomics. 14(5):329–338. doi:10.1093/bfgp/elv028.

Weaver BAA, Cleveland DW. 2006. Does aneuploidy cause cancer? Curr Opin Cell Biol. 18(6):658–667. doi:10.1016/j.ceb.2006.10.002.

Wilhelm T, Said M, Naim V. 2020. DNA Replication Stress and Chromosomal Instability: Dangerous Liaisons. Genes (Basel). 11(6). doi:10.3390/genes11060642.

Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, Amon A. 2008. Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science. 322(5902):703–709. doi:10.1126/science.1160058.

Yang Q, Rasmussen SA, Friedman JM. 2002. Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. Lancet. 359(9311):1019–1025. doi:10.1016/s0140-6736(02)08092-3.

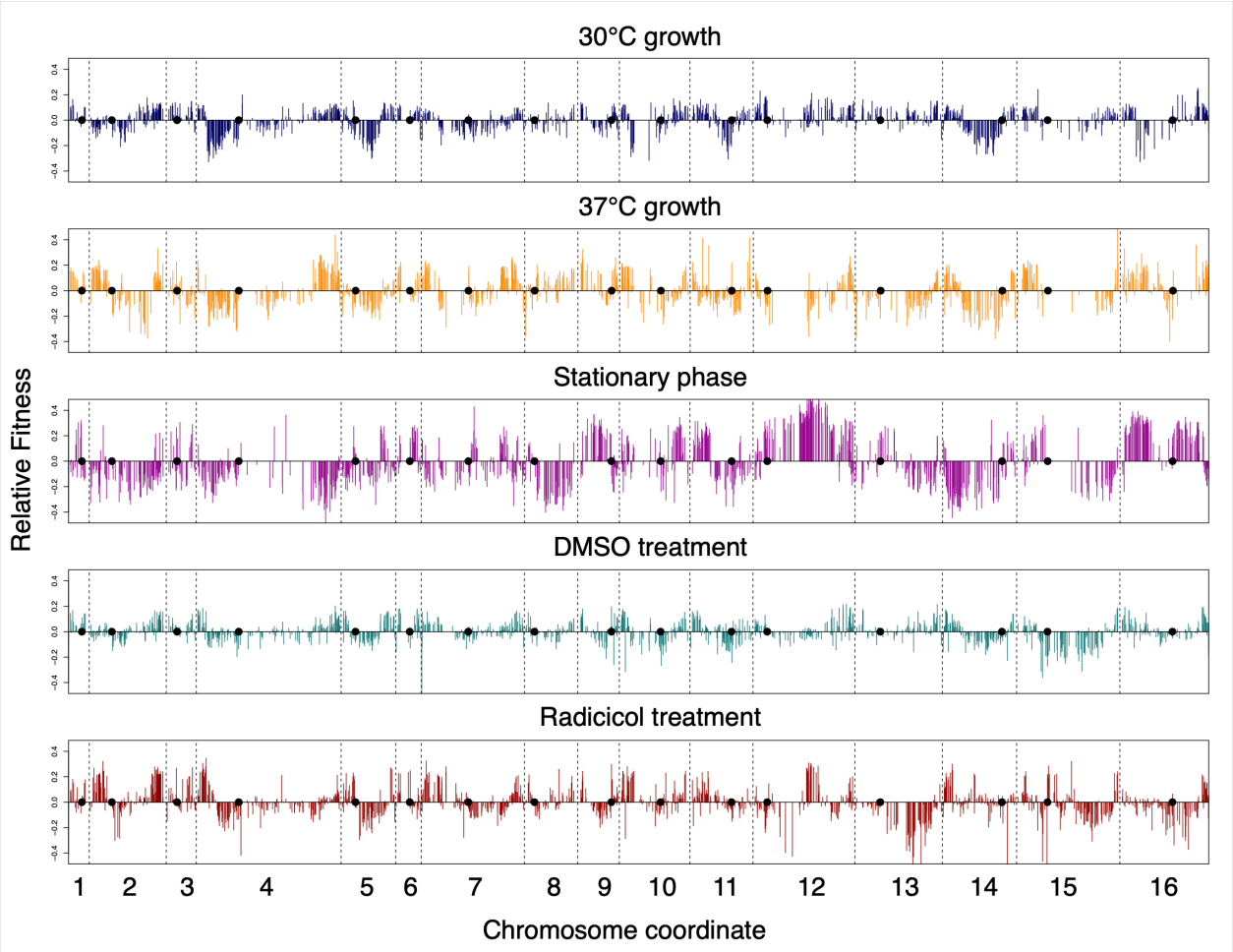
Yurov YB, Vorsanova SG, Demidova IA, Kolotii AD, Soloviev IV, Iourov IY. 2018. Mosaic Brain Aneuploidy in Mental Illnesses: An Association of Low-level Post-zygotic Aneuploidy with Schizophrenia and Comorbid Psychiatric Disorders. Curr Genomics. 19(3):163–172. doi:10.2174/1389202918666170717154340.

Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng C-Z, Wala J, Mermel CH, et al. 2013. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 45(10):1134–1140. doi:10.1038/ng.2760.

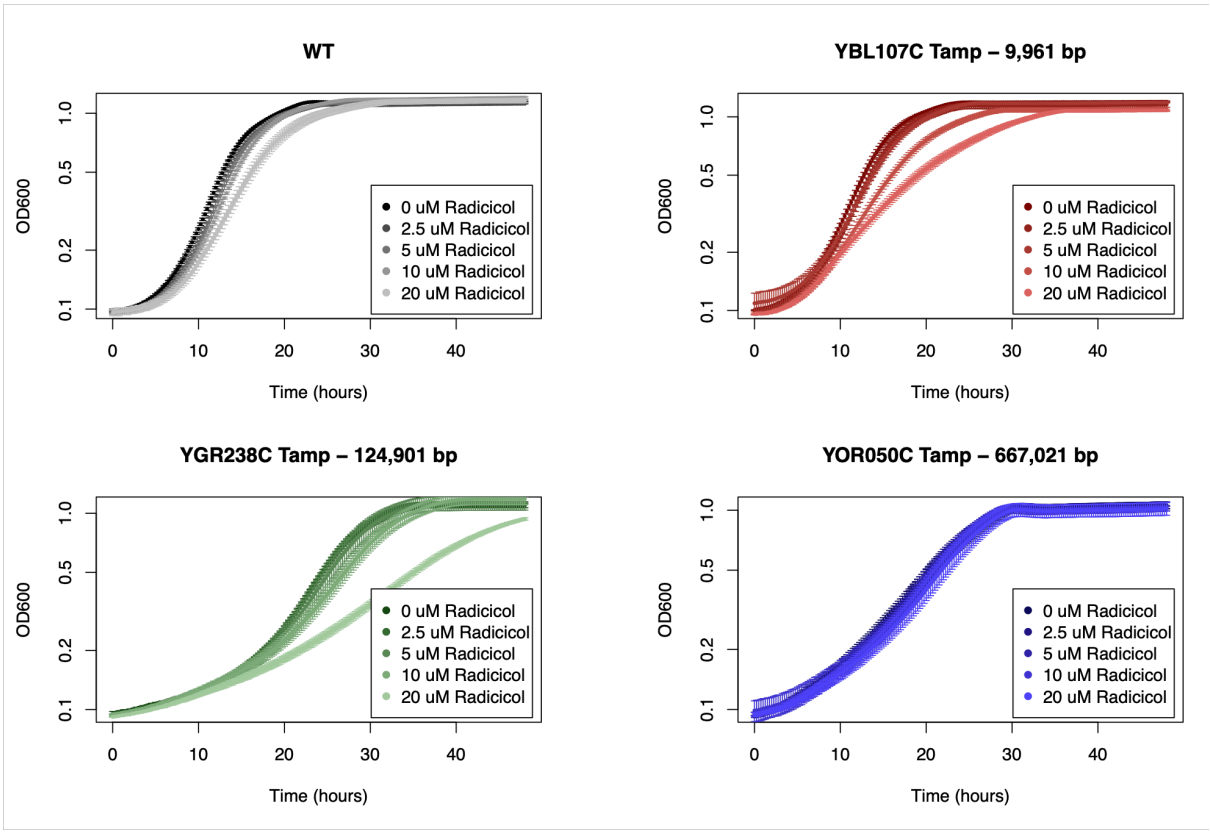
Zarrei M, Burton CL, Engchuan W, Young EJ, Higginbotham EJ, MacDonald JR, Trost B, Chan AJS, Walker S, Lamoureux S, et al. 2019. A large data resource of genomic copy number variation across neurodevelopmental disorders. NPJ Genom Med. 4:26. doi:10.1038/s41525-019-0098-3.

Zhang C-Z, Spektor A, Cornils H, Francis JM, Jackson EK, Liu S, Meyerson M, Pellman D. 2015. Chromothripsis from DNA damage in micronuclei. Nature. 522(7555):179–184. doi:10.1038/nature14493.

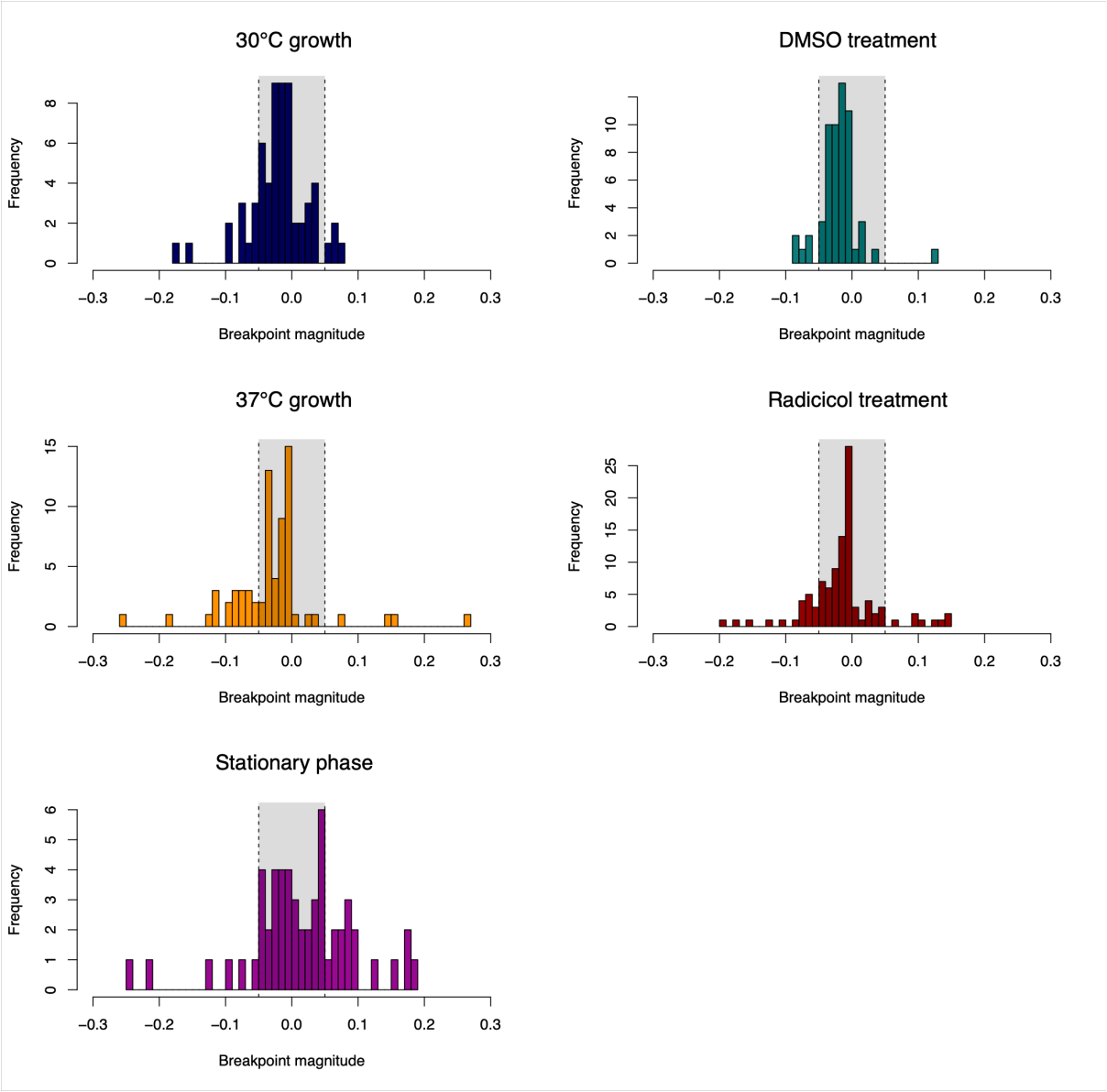
Appendix



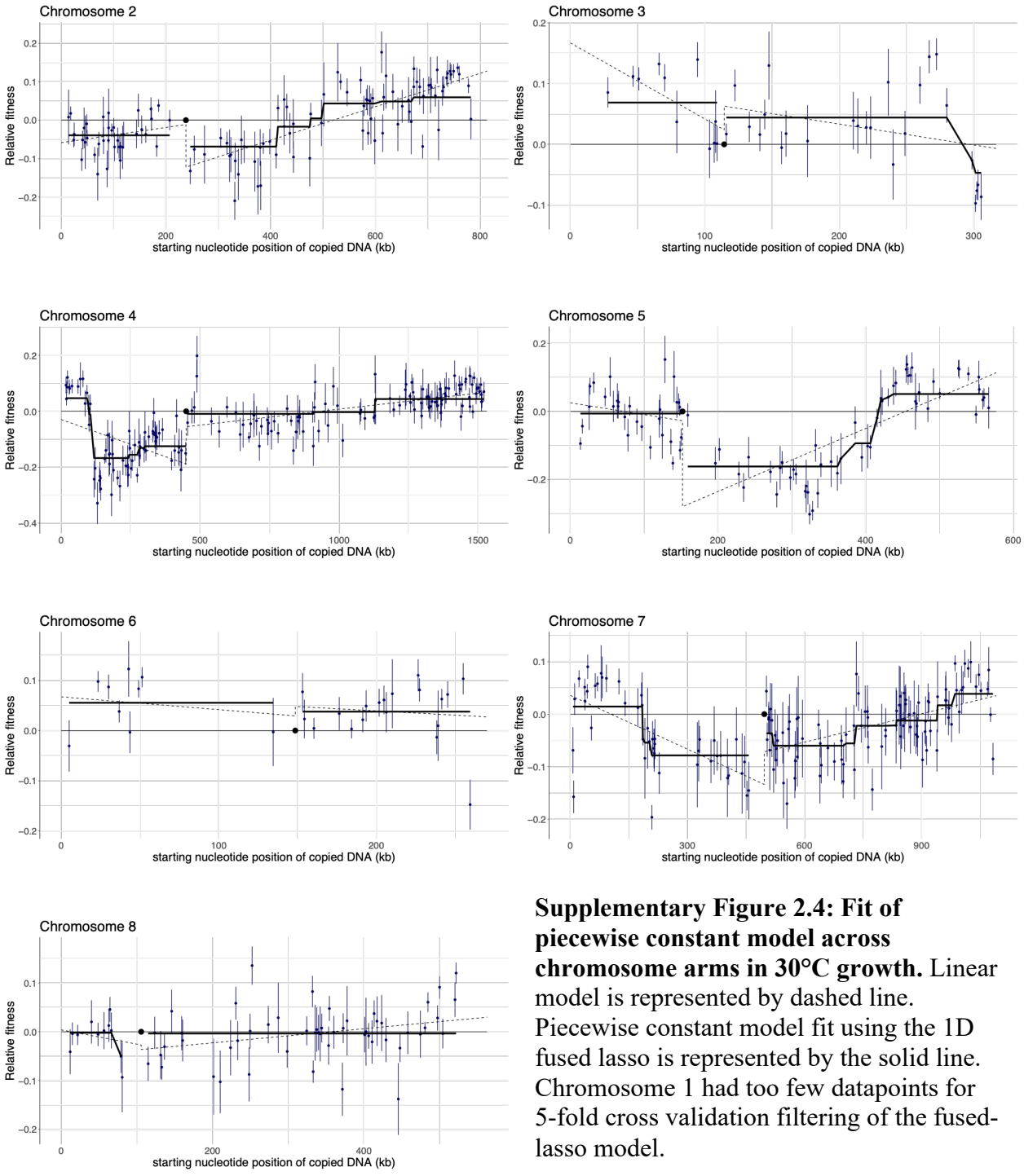
Supplementary Figure 2.1: Fitness of all Tamp strains across the entire genome for each experimental condition. Each bar represents the relative fitness of a Tamp that initiates at that site in the genome and extends to the telomere on the same arm. Relative fitness values are normalized to the mean of all strains for each condition (see Methods). Chromosomes are delimited by dashed lines. Centromeres are represented by black dots.



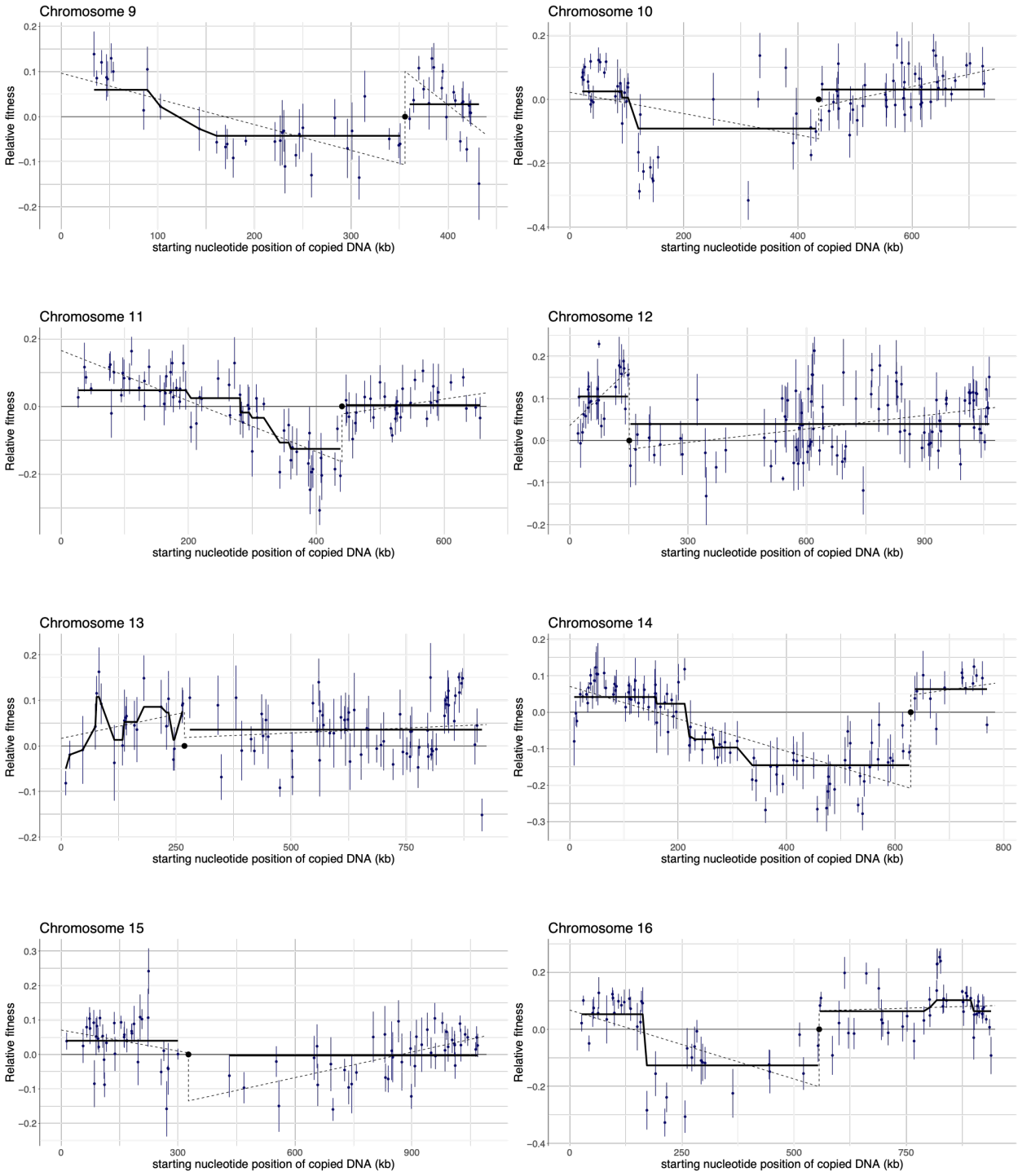
Supplemental Figure 2.2: Effect of radicicol treatment on wild type and Tamp strain growth. Strains were grown in triplicate in a BioTek Synergy H1 Multi-Mode plate reader. Despite YOR050C being the longest Tamp that should be the most sensitive, strains disomic for chromosome 11 do not show sensitivity to Hsp90 inhibitor.



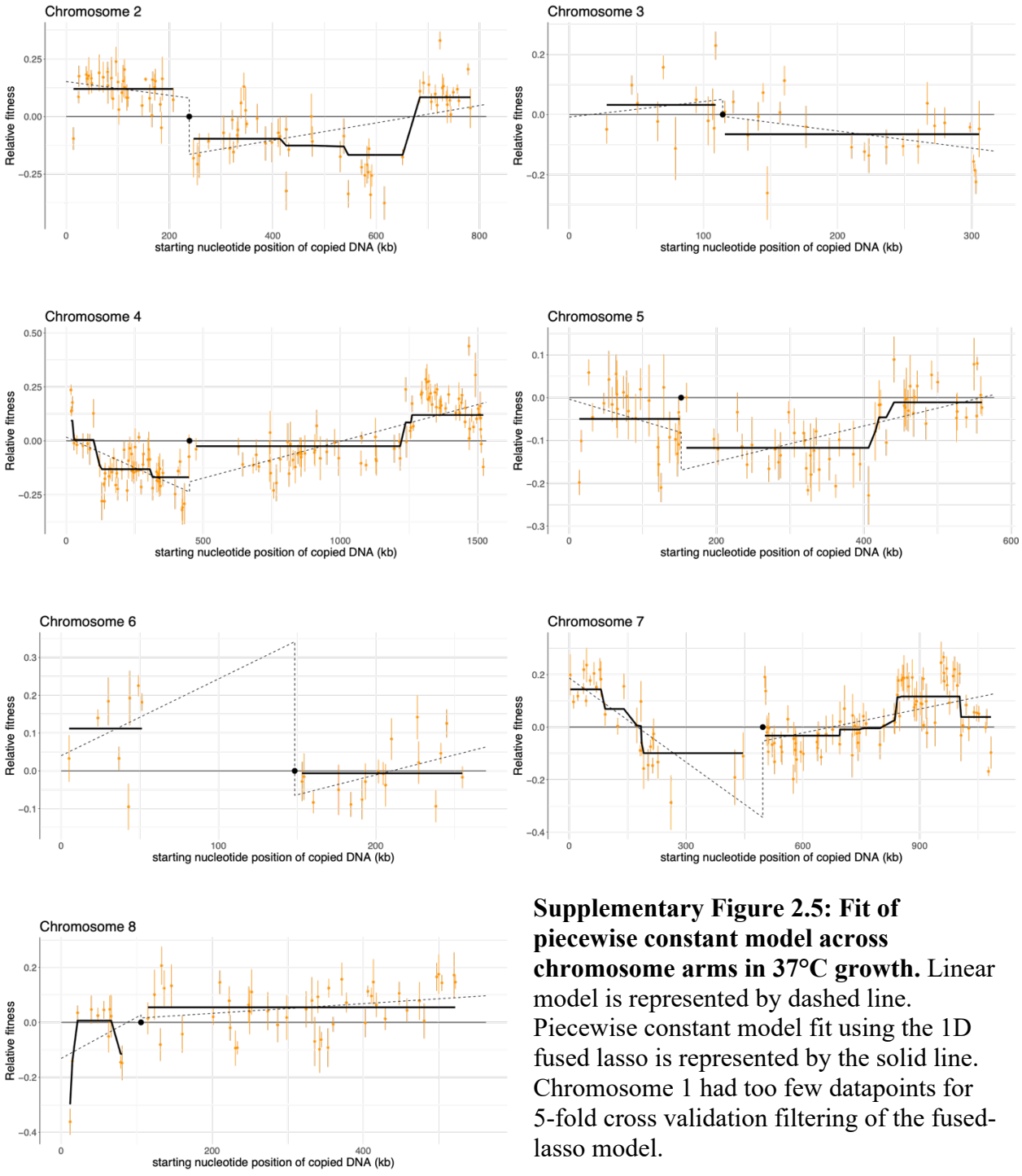
Supplementary Figure 2.3: Breakpoint histograms across all conditions tested. Breakpoint magnitude is calculated as the magnitude of the step points of the piecewise constant model. Upsteps, regions that improve fitness when amplified have a positive magnitude and downsteps and negative magnitude. We narrowed our candidate list of breakpoints as those outside our +/- 0.05 cutoff (dotted lines).



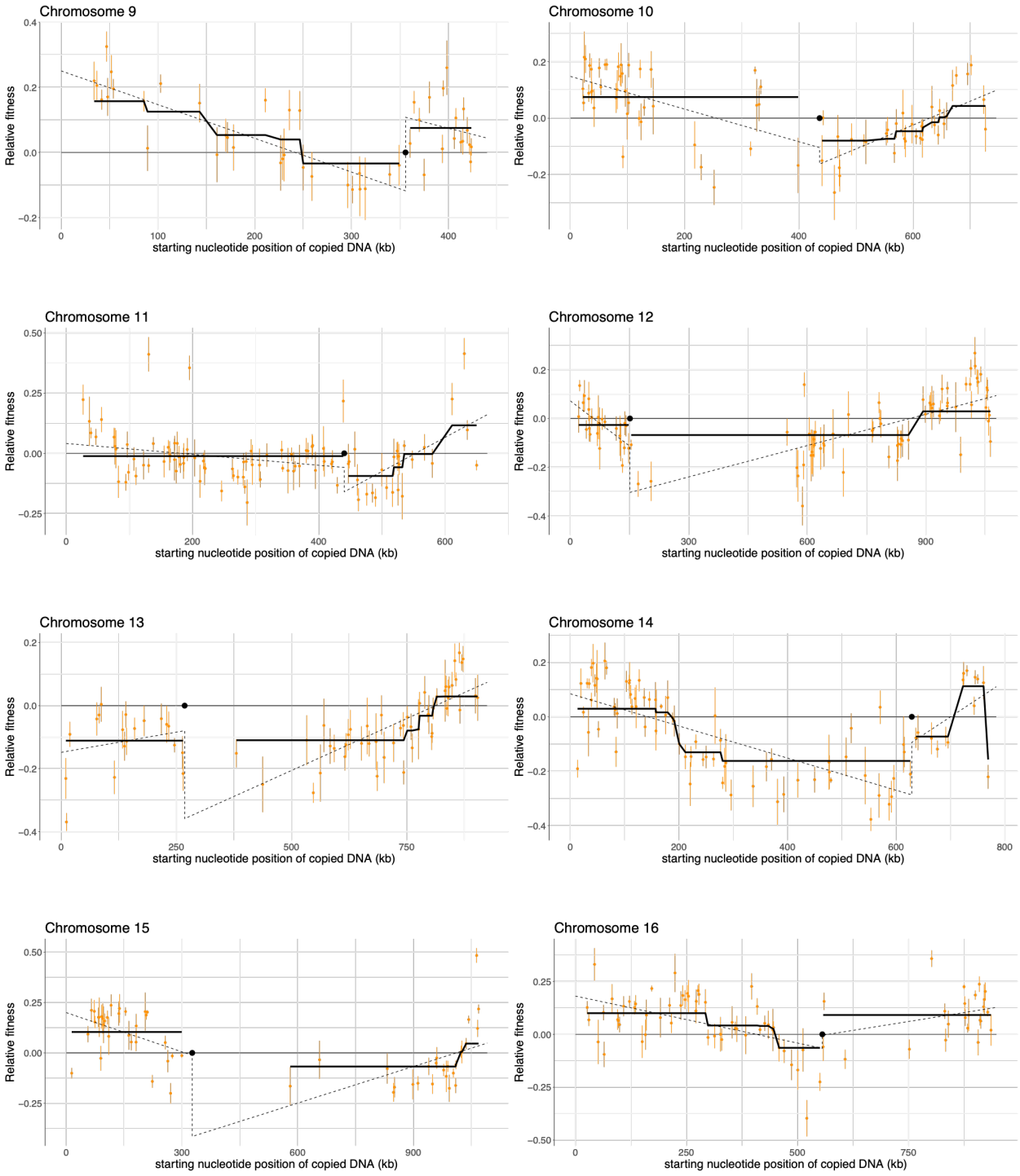
Supplementary Figure 2.4: Fit of piecewise constant model across chromosome arms in 30°C growth. Linear model is represented by dashed line. Piecewise constant model fit using the 1D fused lasso is represented by the solid line. Chromosome 1 had too few datapoints for 5-fold cross validation filtering of the fused-lasso model.



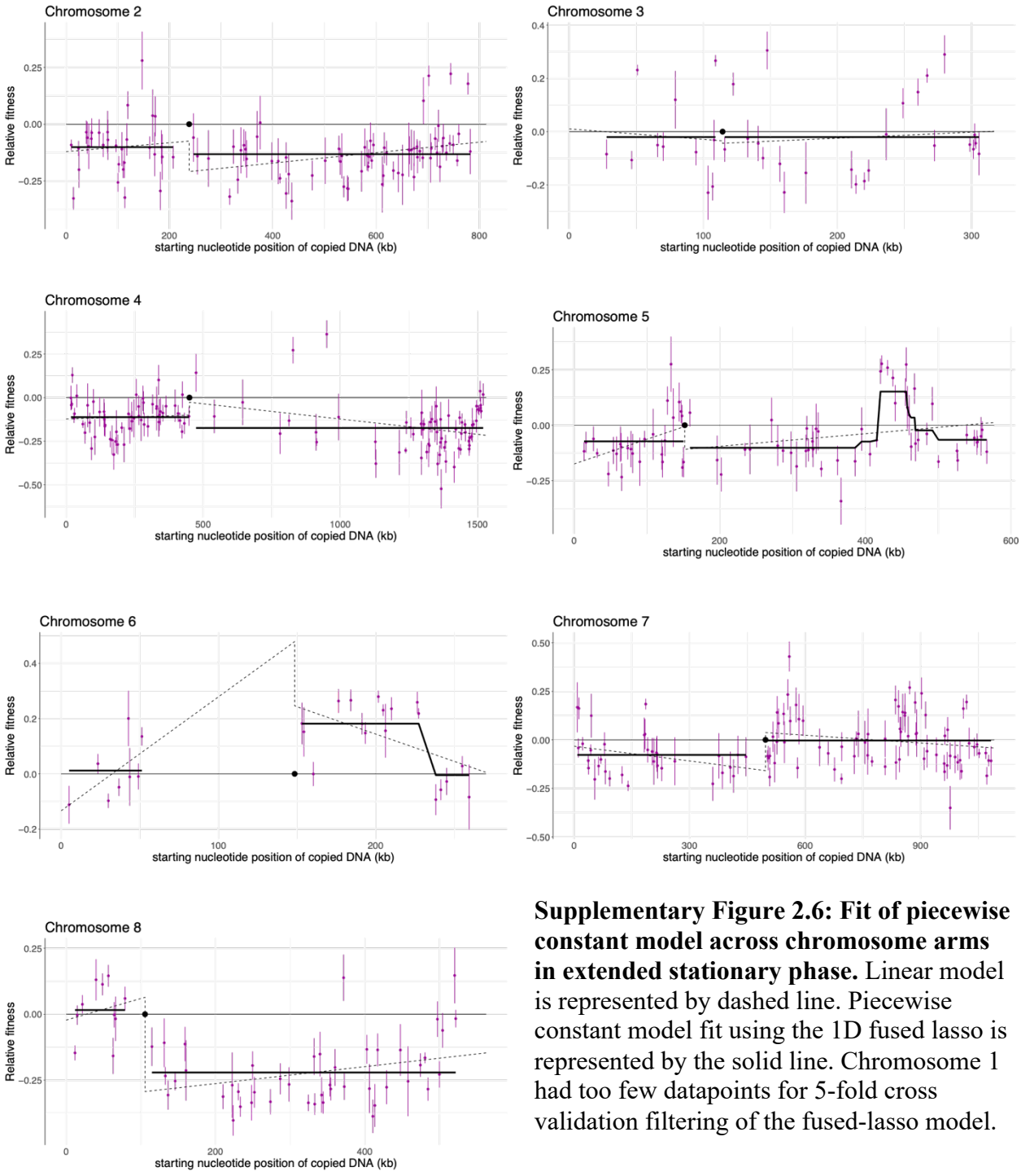
Supplementary Figure 2.4 (continued): Fit of piecewise constant model across chromosome arms in 30°C growth.



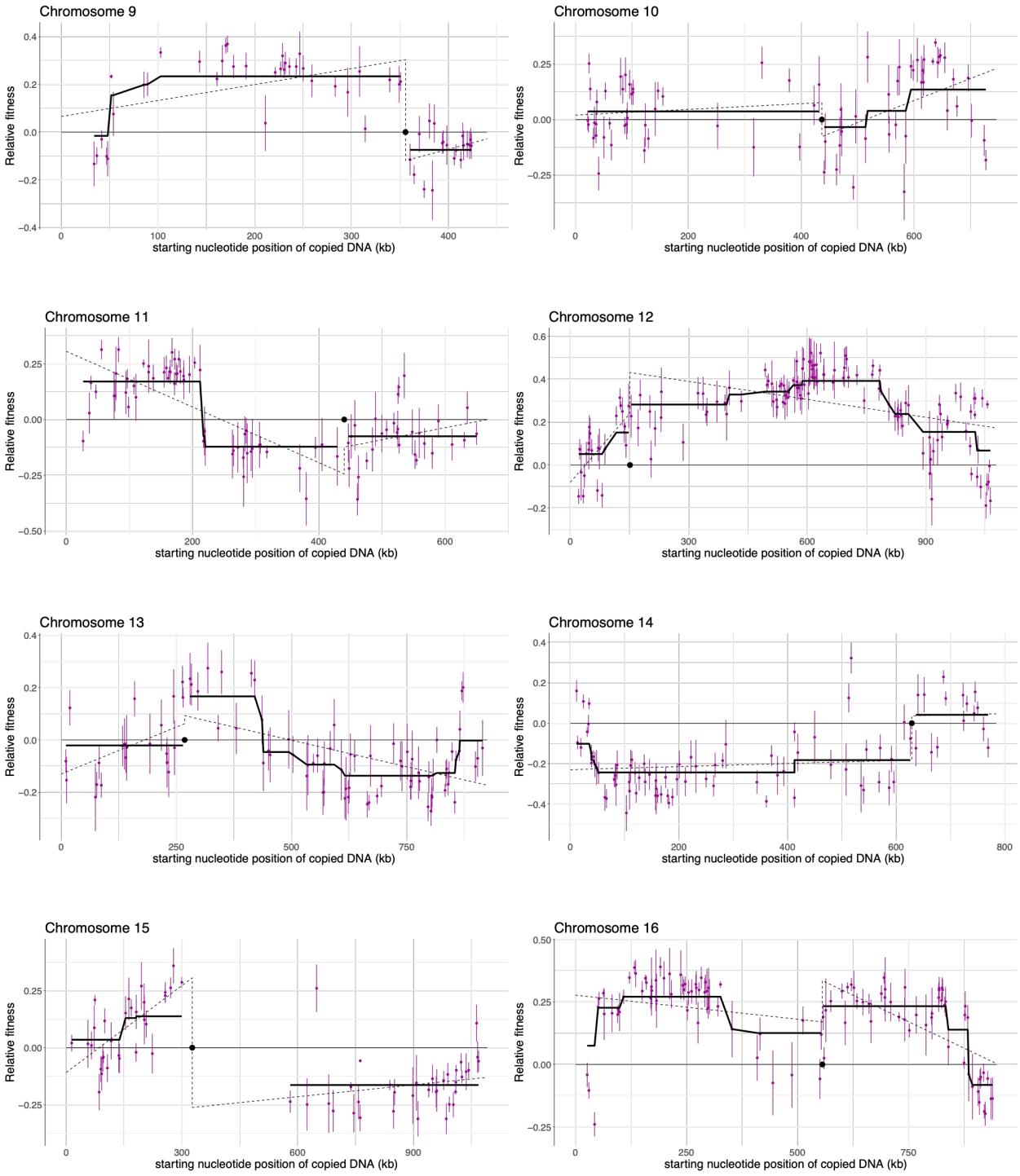
Supplementary Figure 2.5: Fit of piecewise constant model across chromosome arms in 37°C growth. Linear model is represented by dashed line. Piecewise constant model fit using the 1D fused lasso is represented by the solid line. Chromosome 1 had too few datapoints for 5-fold cross validation filtering of the fused-lasso model.



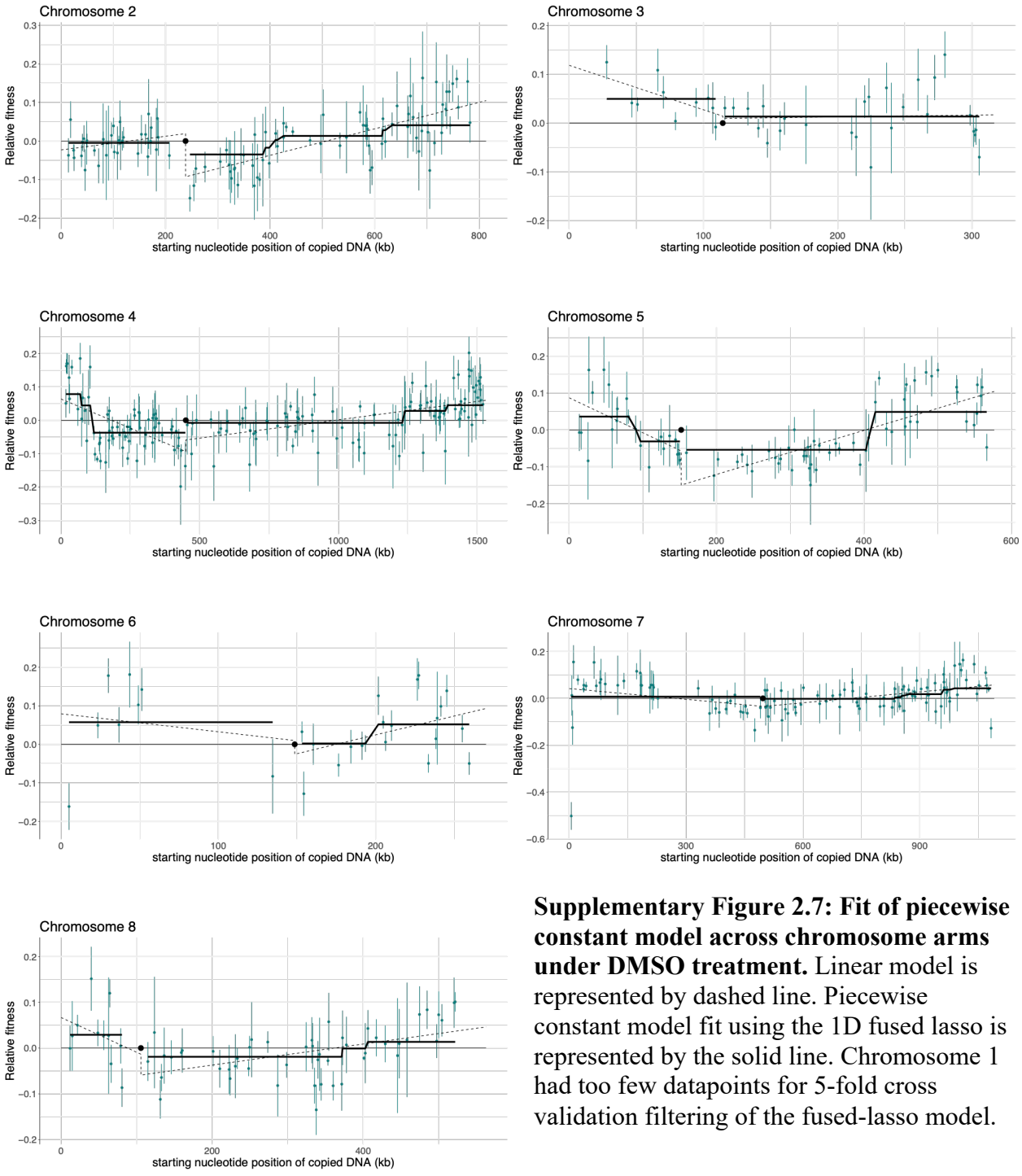
Supplementary Figure 2.5 (continued): Fit of piecewise constant model across chromosome arms in 37°C growth.



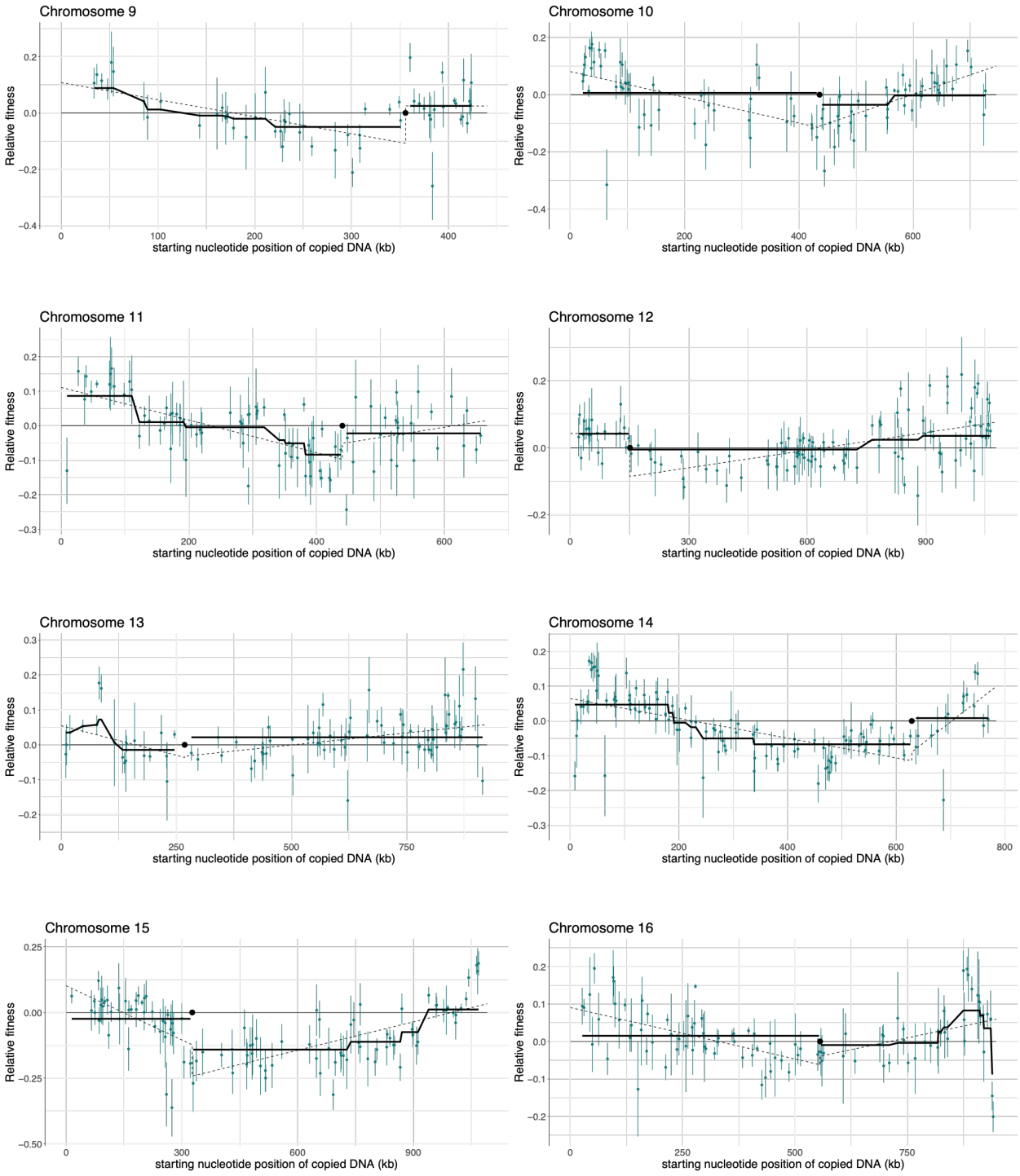
Supplementary Figure 2.6: Fit of piecewise constant model across chromosome arms in extended stationary phase. Linear model is represented by dashed line. Piecewise constant model fit using the 1D fused lasso is represented by the solid line. Chromosome 1 had too few datapoints for 5-fold cross validation filtering of the fused-lasso model.



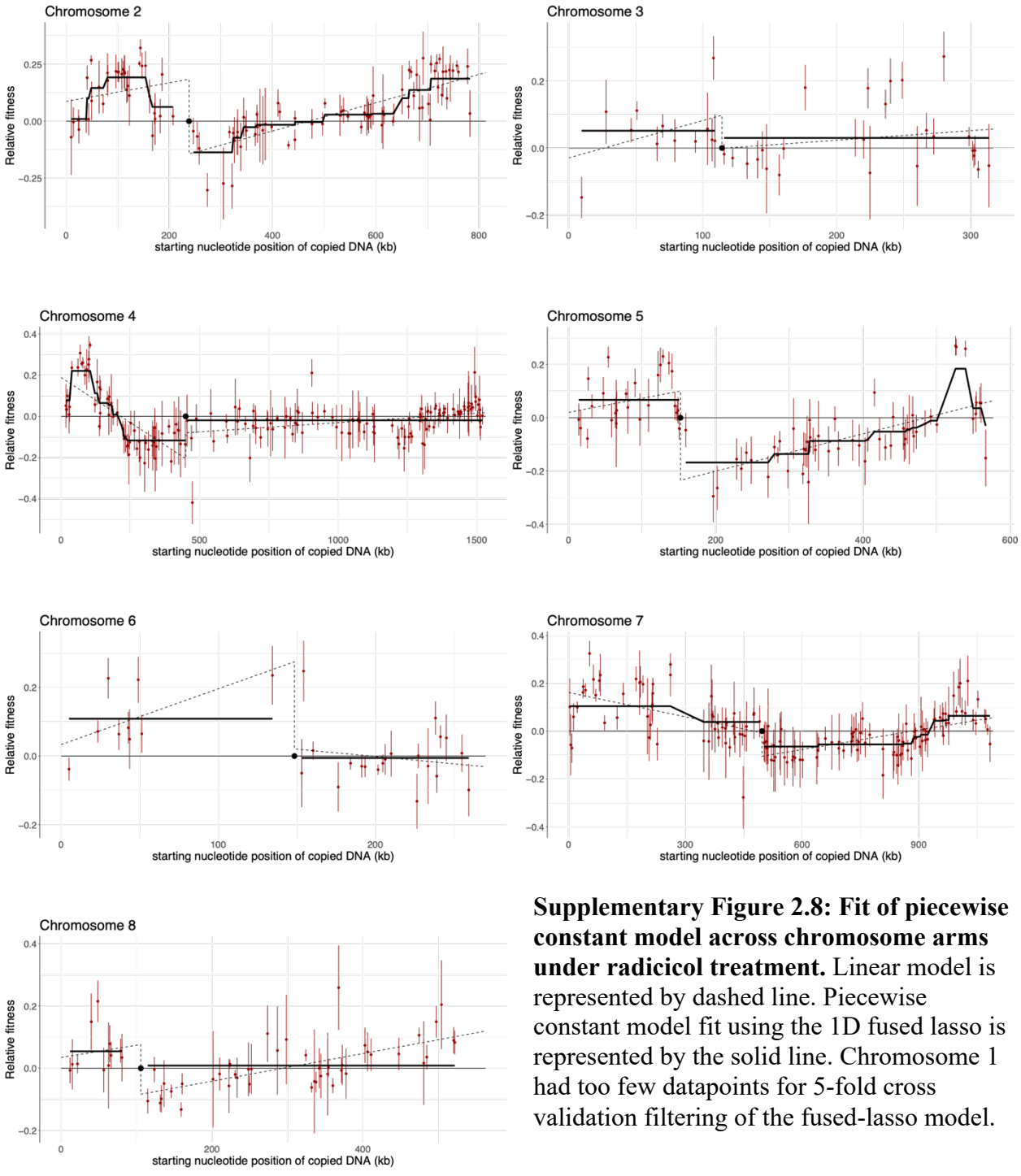
Supplementary Figure 2.6 (continued): Fit of piecewise constant model across chromosome arms in extended stationary phase.



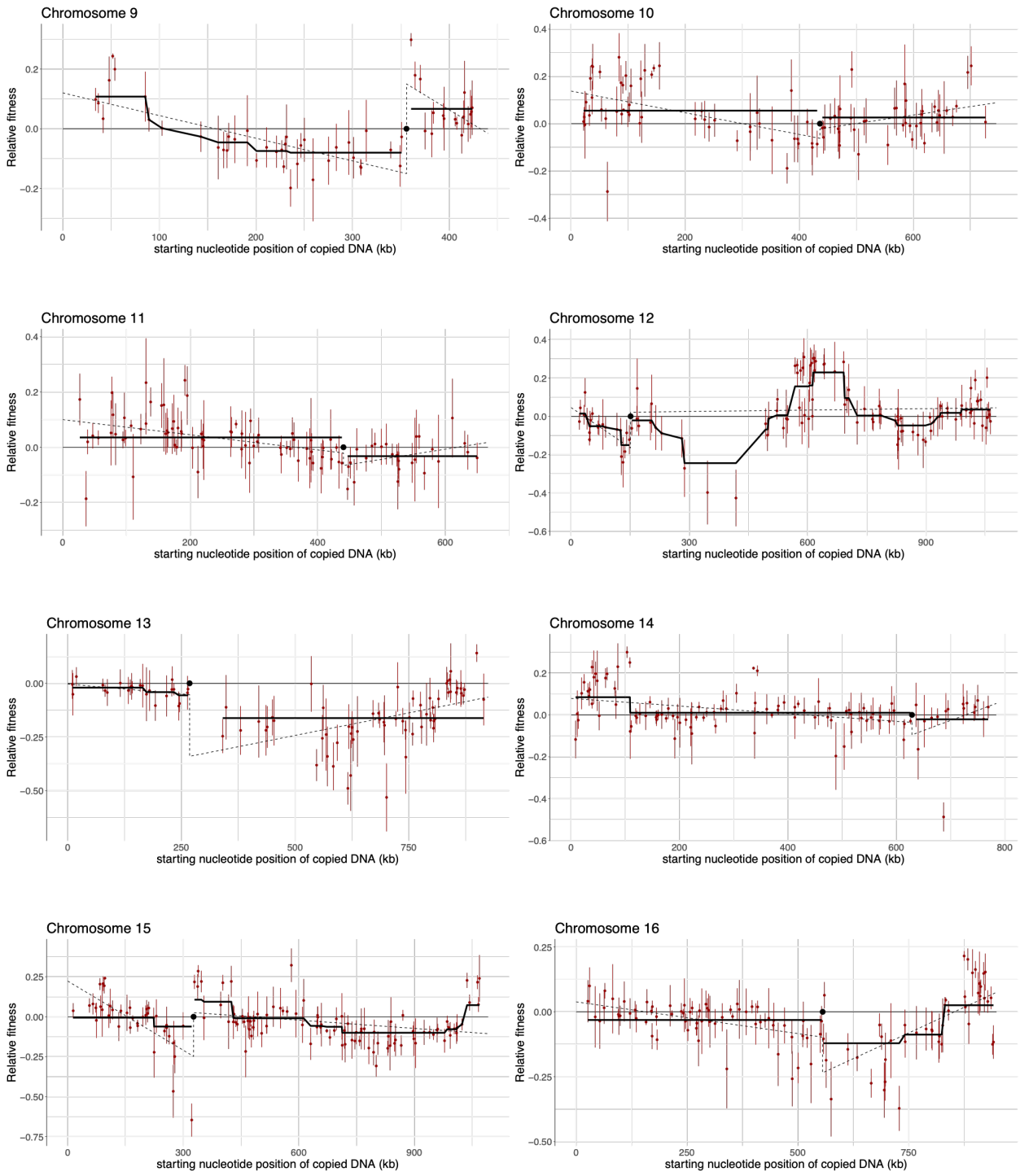
Supplementary Figure 2.7: Fit of piecewise constant model across chromosome arms under DMSO treatment. Linear model is represented by dashed line. Piecewise constant model fit using the 1D fused lasso is represented by the solid line. Chromosome 1 had too few datapoints for 5-fold cross validation filtering of the fused-lasso model.



Supplementary Figure 2.7 (continued): Fit of piecewise constant model across chromosome arms under DMSO treatment.



Supplementary Figure 2.8: Fit of piecewise constant model across chromosome arms under radicicol treatment. Linear model is represented by dashed line. Piecewise constant model fit using the 1D fused lasso is represented by the solid line. Chromosome 1 had too few datapoints for 5-fold cross validation filtering of the fused-lasso model.



Supplementary Figure 2.8 (continued): Fit of piecewise constant model across chromosome arms under radicicol treatment.

Supplementary Table 2.1: Strains used in this study.

Strain name	Strain ID	Strain background	Reference
YOR050C_Tamp	MD361-8F	BY4743 (S288C)	Sunshine, 2015
YGR238C_Tamp	MD361-12E	BY4743 (S288C)	Sunshine, 2015
YBL107C_Tamp	MD362-10C	BY4743 (S288C)	Sunshine, 2015
wlcr Tamp pool	Pool M	BY4743 (S288C)	Sunshine, 2015
wrcl Tamp pool	Pool N	BY4743 (S288C)	Sunshine, 2015
Yeast knockout diploid wlcr genes	Pool J	BY4743 (S288C)	Sunshine, 2015, Tong et al. 2004
Yeast knockout diploid wrcl genes	Pool K	BY4743 (S288C)	Sunshine, 2015, Tong et al. 2004

Supplementary Table 2.2: Primers used in this study.

Primer ID	Primer name	Primer Sequence	Notes
AK1	SA_3_PCR_R1	AATGATACGGCGACCACCGAGAT CTACACGGCCGTCGACGAAAAGT GCCACCTGACGTC	Amplify Tamp barcodes for sequencing - Reverse primer for wlcr and wrcl pools
AK2	SA_3_PCR_F#	CAAGCAGAAGACGGCATACGAG AT NNNNNN GCGCTCCGAGCGGATGTCCACGA GGTCTCT	Amplify Tamp barcodes for sequencing - Reverse primer for wlcr pool
AK3	SA_4_PCR_F#	CAAGCAGAAGACGGCATACGAG AT NNNNNN GCGCTCCGAGCGCGGTGTCGGTC TCGTAG	Amplify Tamp barcodes for sequencing - Reverse primer for wrcl pool
AK4	SA_2_SEQ_index	CTACGAGACCGACACCGCGCTCG GAGC	Illumina sequencing of wrcl pool - index
AK5	SA_2_SEQ_R1	GCGCTCCGAGCGCGGTGTCGGTC TCGTAG	Illumina sequencing of wrcl pool - read 1
AK6	SA_3_SEQ_F1	ACACGGCCGTCGACGAAAAGTGC CACCTGACGTC	Illumina sequencing of wrcl, wlcr pools - read 2
AK7	SA_SEQ_index	AGAGACCTCGTGGACATCCGCTC GGAGCGC	Illumina sequencing of wlcr pool - index
AK8	UP BAR F#	CAAGCAGAAGACGGCATACGAG AT NNNNNN GACCTTACGCGTGGATGTCCACG AGGTCTCT	Amplify deletion collection Uptag barcodes - forward primer
AK9	UP BAR R1	AATGATACGGCGACCACCGAGAT CTACAC AGAGGCCGTCACGTCGACCTGCA GCGTACG	Amplify deletion collection Uptag barcodes - reverse primer
AK10	UP SEQ F1	AGA GGC CGT CAC GTC GAC CTG CAG CGT ACG	Illumina sequencing of Uptag barcodes from deletion collection pool - read1
AK11	UP SEQ index	AGAGACCTCGTGGACATCC ACGCGTAAGGTC	Illumina sequencing of Uptag barcodes from deletion collection pool - index1

Supplementary Table 2.3: Deletion strains excluded from analysis. Strains were excluded if the relative fitness of the deletion strain was not within +/- 0.05 in a particular experimental condition.

Gene deletion	Condition	Fitness
YAL038W	DMSO	0.07320438
YBR074W	DMSO	0.05154026
YBR214W	DMSO	0.0746441
YGL024W	DMSO	0.03970256
YJL109C	DMSO	0.04005687
YMR061W	DMSO	0.05052679
YMR151W	DMSO	0.0511167
YMR203W	DMSO	0.03482782
YNL339C	DMSO	0.08419097
YOR207C	DMSO	0.03663559
YPR033C	DMSO	0.03684959
YPR087W	DMSO	0.06371675
YAL038W	Radicicol	0.06698526
YBR074W	Radicicol	0.0480297
YBR214W	Radicicol	0.07742095
YGL024W	Radicicol	0.05132994
YJL109C	Radicicol	0.05184295
YLR326W	Radicicol	0.03051983
YMR061W	Radicicol	0.07686722
YMR106C	Radicicol	0.05257769
YMR151W	Radicicol	0.06296331
YMR203W	Radicicol	0.09809153
YNL339C	Radicicol	0.09635215
YOR207C	Radicicol	0.08221812
YPR033C	Radicicol	0.04986137
YPR087W	Radicicol	0.05810946
YGR260W	Stationary	0.04490344

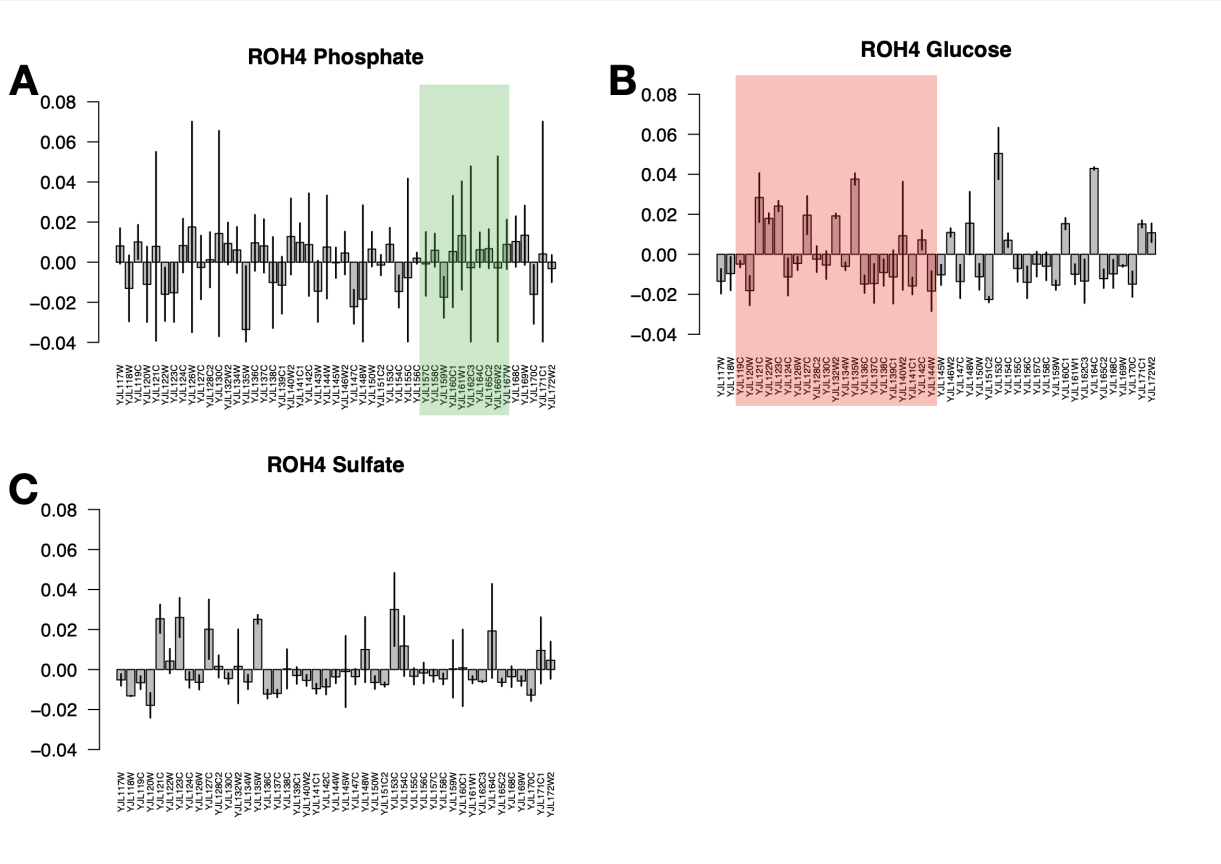
Supplementary Table 2.4: Genes present at candidate fitness breakpoints.

Condition	Chr.	Start coordinate	Stop coordinate	Magnitude	Step type	Genes
Radicicol	2L	39145	40831	0.0902	upstep	*BNA4*
Radicicol	2L	153848	161699	-0.0832	downstep	*HEK2; RIB1; STU1*
Radicicol	2R	320416	321876	0.0642	downstep	*CST26*
T30	2R	409169	411054	0.0518	downstep	*TEC1*
Radicicol	2R	633622	650368	0.0646	downstep	*KTR3;* YBR206W*; FTH1; DUR1,2; YBR209W; ERV15; AME1; NGR1*
T37	2R	650368	683428	0.2507	downstep	*MET8; SDS24; HPC2; YBP1; ATG12; PYC2;* YBR219C; YBR220C*; PDB1;* YBR221W-A*; PCS60; TDP1;* YBR223W-A*;* YBR224W*;* YBR225W*;* YBR226C*; MCX1; SLX1; ROT2; OM14;* YBR230W-A*; SWC5;* YBR232C
T30	3R	276764	297049	-0.0704	upstep	*MSH3; CDC39; CDC50; OCA4;* YCR095W-A*; HMRA2; HMRA1;* YCR097W-A
T37	4L	22823	28985	-0.0894	downstep	*LRG1;* YDL240C-A*; ADY3*
Radicicol	4L	30657	38867	0.1425	upstep	*AIM6; PHO13; YPD1; GYP7; MFG1; OST4*
T37	4L	99561	122216	-0.1177	downstep	*TRM8; MGT1;* YDL199C*; GGC1; ASF2;* YDL196W*; SEC31; SNF3; NUS1; ARF1; RPL35A; UFD2*
DMSO	4L	104917	118707	-0.0816	downstep	*ASF2;* YDL196W*; SEC31; SNF3; NUS1; ARF1; RPL35A*
Radicicol	4L	104917	122216	-0.1087	downstep	*ASF2;* YDL196W*; SEC31; SNF3; NUS1; ARF1; RPL35A; UFD2*
T30	4L	104917	118707	-0.1566	downstep	*ASF2;* YDL196W*; SEC31; SNF3; NUS1; ARF1; RPL35A*
Radicicol	4L	203039	222427	-0.0685	downstep	*BPL1; RPO21; SCM3; RGT2; ARF2; RPL35B; RDI1; PPH21*
T37	4R	1215016	1236558	0.1103	downstep	*XRS2; DXO1;* YDR371C-A*; CTS2; VPS74; FRQ1; PHO92; WIP1; BCS1; ARH1; ATP17; LSM6; SDH6; RGA2; ARO10;* YDR381C-A
DMSO	5R	404813	414481	0.0699	downstep	*YCK3; DSE1; RSP5; NSA2*
T30	5R	404813	414481	0.0746	downstep	*YCK3; DSE1; RSP5; NSA2*
Stat	5R	414481	417281	0.2192	downstep	*LCP5; VFA1*
T30	5R	414481	417281	0.0537	downstep	*LCP5; VFA1*
Stat	5R	455146	456319	-0.0714	upstep	*MAG1*
Stat	5R	466208	468370	-0.0573	upstep	*PEA2*
Radicicol	5R	499347	523369	0.1953	downstep	*SPT2; RAD4; GCG1; CHD1;* YER165C-A; *PAB1; DNF1; BCK2; CCA1*
Radicicol	5R	538434	548421	-0.1484	upstep	*GRX4; TMT1;*YER175W-A *; ECM32; BMH1; PDA1*
Radicicol	5R	559454	566230	-0.0661	upstep	*PUG1;* YER186C
Stat	6R	226963	237368	-0.1863	upstep	*CDC26;*YFR036W-A *; RSC8; IRC5; OSW7; SAP155*
T37	7L	81426	92512	-0.0747	downstep	*NIF3; FRA2; MDM34;* YGL218W*;* YGL217C*; KIP3; CLG1;* YGL214W*; SKI8; VAM7*
T37	7L	140368	173285	-0.0623	downstep	*HOS2;* YGL193C*; IME4; COX13; CDC55; RPS26A;* YGL188C-A*;* YGL188C*; COX4; TPN1;* YGL185C*;*

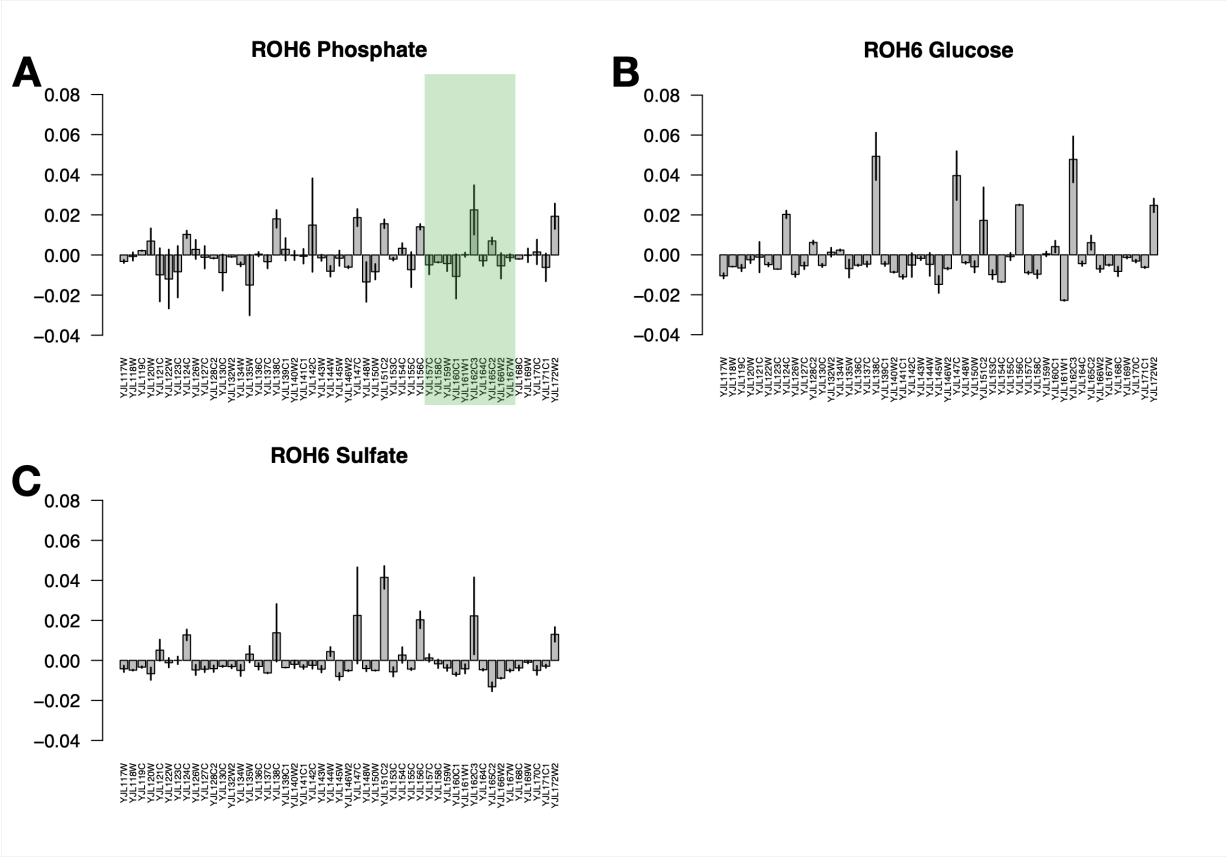
						STR3; MND1; YGL182C*; GTS1; ATG1; TOS3; MPT5;* YGL177W*;* YGL176C
T30	7L	184153	186059	-0.0623	downstep	*SPO74*
T37	7L	184153	186059	-0.0644	downstep	*SPO74*
Radicicol	7L	261648	347119	-0.0659	downstep	*SNT2; CEG1; RSM23; CWC23; SOH1; SCS3; MET13; MON1; RPS2;* YGL123C-A*; NAB2; GPG1; PRP43; COQ8;* YGL118C; YGL117W*; CDC20; SNF4;* YGL114W*; SLD3; TAF6; NSA1; CUE3;* YGL109W*;* YGL108C*; RMD9; MLC1; ARC1; VPS73; RPL28;* YGL102C*;* YGL101W*; SEH1; LSG1; USE1; SRM1; TOS8; VPS45; PAN2; SPC105; NUP145; NBP35; LIF1; MF(ALPHA)2;* YGL088W*; MMS2*
T37	7R	834689	840824	0.0884	downstep	*PUS6; LSO2; PSD2*
Radicicol	7R	922175	937124	0.0579	downstep	*RSM27; GPI1; CCH1; CRM1;* YGR219W*; MRPL9*
T37	7R	1000927	1004624	-0.0782	upstep	*ENO1; COQ6*
T37	8L	11923	14901	0.1576	upstep	*PAU13;* YHL045W; YHL044W
T37	8L	14901	21783	0.1474	upstep	*ECM34;* YHL042W; YHL041W*; ARN1*
T37	8L	66242	78936	-0.1213	downstep	*OPI1;* YHL019W-A*; APM2;* YHL018W*;* YHL017W*; DUR3;* YHL015W-A*; RPS20; YLF2; OTU2*
Stat	9L	48091	51759	0.1702	upstep	*UBP7*
Radicicol	9L	85366	89230	-0.0771	downstep	*AXL2; REV7*
T37	9L	142928	160887	-0.0718	downstep	*HIS5;* YIL115W-A*; NUP159; POR2; SDP1; HOS4; COX5B; HPM1; SEC24*
T37	9L	246392	249991	-0.0737	downstep	YIL060W*;* YIL059C*;* YIL058W*; RGI2*
T30	10L	102227	120449	-0.0970	downstep	*SET2; ERG20; QCR8; HAL5; TPK1;* YJL163C*; JJJ2; FMP33; PIR5*
Stat	10R	510239	516461	0.0731	downstep	*URB2; NUP85*
Stat	10R	583733	590872	0.0960	downstep	YJR084W*;* YJR085C*; STE18;* YJR087W*; EMC2; BIR1*
DMSO	11L	110464	122517	-0.0760	downstep	*COY1; STE3;* YKL177W*; LST4; ZRT3; TPO5*
Stat	11L	212043	217344	-0.2447	downstep	YKL123W*; SRP21; DGR2*
T30	11L	317764	342852	-0.0733	downstep	*MNR2;* YKL063C*; MSN4; BLI1; FBA1; MPE1; TOA2; NUP120; TMA19; OAR1; DEF1;* YKL053W*; MDM35; ASK1; SFK1*
T37	11R	530946	534067	0.0552	downstep	YKR051W*; MRS4*
T37	11R	579209	608946	0.1192	downstep	*AIM29*; YKR075C*;* YKR075W-A*; ECM4; MSA2;* YKR078W*; TRZ1; MTD1; RPF2; NUP133; DAD2; HBS1; MRPL20; PRP16; OMA1; TVP38; TGL4*
Stat	12L	87403	116432	0.0776	upstep	*ISA1; HSP104; PAU17; SSA2; POM33; HIF1; SPA2;* YLL020C*;* YLL019W-A*; KNS1; COX19; DPS1*
Radicicol	12L	125534	127523	-0.0773	downstep	*YEH1*
Radicicol	12R	278862	286820	-0.1290	upstep	*LAM6; RFU1; BUD20; RPL10;* YLR076C*; FMP25; BOS1*
Radicicol	12R	417006	491866	0.1754	downstep	*RKM5; NHA1; SLS1;* YLR140W*; RRN5; PUT1; DPH6; ACF2; RMP1; SPE4;* YLR146W-A*; SMD3; PEP3;* YLR149C; YLR149C-A*; STM1; PCD1;* YLR152C*; ACS2; RNH203;* YLR154C-G*;* YLR154C-H*;* YLR154W-A*;* YLR154W-B*; TAR1;*

						YLR154W-E; YLR154W-F*; ASP3-1;* YLR156C-A; YLR156W*; ASP3-2;* YLR157C-C*;* YLR157W-D*;* YLR157W-E*; ASP3-3;* YLR159C-A*;* YLR159W*; ASP3-4;* YLR161W*;* YLR162W*; RRT15*
Radicicol	12R	495429	500733	0.0572	downstep	*SEC10; RPS31; UPS2;* YLR169W*; APS1*
Radicicol	12R	547853	566654	0.1504	downstep	*PBA1; YKE2; COQ9; ; MSS51; QRI5; HMX1; ENT2; HRD3; SEC13; PNP1; CLB4; ATG38; TUB4;*
Radicicol	12R	611794	612367	0.0680	downstep	YLR236C
Radicicol	12R	689083	691555	-0.1339	upstep	*PIG1*
Radicicol	12R	699999	724721	-0.0913	upstep	YLR278C; YLR279W; YLR280C*;* YLR281C; YLR282C; YLR283W*; ECI1;* YLR285C-A*; NNT1; CTS1;* YLR286W-A*;* YLR287C*; RPS30A; MEC3; GUF1; COQ11; GCD7; SEC72; GSP1;* YLR294C*; ATP14;* YLR296W; YLR297W
Stat	12R	784913	805784	-0.0624	upstep	*NMA1; REC102; CHS5; JIP3; MID2; RPS25B;* YLR334C*; NUP2; SGD1; VRP1; OPI9*
Stat	12R	855538	888851	-0.0836	upstep	YLR366W*; RPS22B; MDM30; SSQ1; ARC18; ROM2; ELO3; VID22;* YLR374C*; STP3; PSY3; FBP1; SEC61;* YLR379W*; CSR1; CTF3; NAM2; SMC6*
T37	12R	855538	888851	0.0971	downstep	YLR366W*; RPS22B; MDM30; SSQ1; ARC18; ROM2; ELO3; VID22;* YLR374C*; STP3; PSY3; FBP1; SEC61;* YLR379W*; CSR1; CTF3; NAM2; SMC6*
Stat	12R	1024189	1028854	-0.0869	upstep	*GMC2;* YLR446W*; VMA6*
T30	13L	46942	74398	0.0513	upstep	*BUL2; COQ5; ZDS2;* YML108W*; PML39; URA5; SEC65; MDM1; NUP188; CAC2;* YML101C-A*; CUE4;* YML100W-A*; TSL1;* YML099W-A
T30	13L	74398	77267	0.0651	upstep	*TSL1;* YML099W-A*; ARG81*
DMSO	13L	87123	115734	-0.0661	downstep	*RPM2;* YML090W*;* YML089C*; UFO1; AIM33; ALO1; TUB1;* YML084W*;* YML083C; YML082W*; TDA9; ATP18; DUS1;* YML079W*; CPR3; BET5; WAR1*
T30	13L	87123	115734	-0.0793	downstep	*RPM2;* YML090W*;* YML089C*; UFO1; AIM33; ALO1; TUB1*; YML084W*;* YML083C; YML082W*; TDA9; ATP18; DUS1;* YML079W*; CPR3; BET5; WAR1*
T30	13L	233457	244149	-0.0600	downstep	*OST6;* YML018C*; PSP2; PPZ1; TAF11; TRM9*
T30	13L	246116	263483	0.0620	upstep	*ERV25;* YML012C-A*; RAD33; SPT5;* YML009W-B*;* YML009C-A*; MRPL39; ERG6; YAP1;* YML007C-A*; GIS4; TRM12; GLO1*
Stat	13R	416196	434788	-0.0892	upstep	*PDS5; VPS20; CTF18; SEC14; NAM7; ISF1;* YMR082C
Stat	13R	434788	437491	-0.1233	upstep	*ADH3;* YMR084W
Stat	13R	852630	855399	0.0500	downstep	*TDA1; GOT1*
Radicicol	14L	108467	109095	-0.0743	downstep	*HCH1*
T37	14L	191324	200569	-0.0842	downstep	*ATG2; ZWF1; NAR1*
T30	14L	211922	220645	-0.0901	downstep	*BNI4; CSL4; PDR16; ELA1; URE2*
Stat	14L	412684	412771	0.0609	upstep	YNL114C

T37	14R	695052	721120	0.1851	downstep	*RSM19; DBP6; ZRG17;* YNR040W*; COQ2;* YNR042W*; MVD1; AGA1; PET494; TRM112; FPK1;* YNR048W*; MSO1; LYS9; BRE5; POP2*
T37	14R	760067	765375	-0.2694	upstep	YNR068C*; BSC5*
Stat	15L	138107	153912	0.0950	upstep	*COQ3; HMI1; RFC4; TRM10; YPQ1; SPO21; MSH2; HAL9*
Radicicol	15L	223268	224420	-0.0556	downstep	*GPM3*
Radicicol	15R	424270	427833	-0.1013	upstep	YOR050C*; ETT1; TMC1;* YOR053W
DMSO	15R	909343	936731	0.0861	downstep	*FAA1;* YOR318C*; HSH49; GNT1; PMT3; LDB19; PRO2; FRT1;* YOR325W*; MYO2; SNC2; PDR10*
T37	15R	1009224	1020222	0.0642	downstep	*SNX3; HAP5; VTS1; PDE2; PRT1; PRE10*
Radicicol	15R	1020222	1034807	0.1203	downstep	*PIP2;* YOR364W; YOR365C*;* YOR366W*; SCP1; RAD17; RPS12; MRS6; GPB1*
Stat	16L	43283	51244	0.1520	upstep	YPL264C*; KEL3; FUM1;* YPL261C*;* YPL260W
T30	16L	162632	172033	-0.1792	downstep	*PGC1;* YPL205C*; HRR25; TPK2; AFT2; YIG1; CSM4*
T37	16L	292817	298571	-0.0583	downstep	*GIP3;* YPL136W*; ISU1;* YPL135C-A
Stat	16L	326628	352863	-0.1296	downstep	*MRP51; IDI1; HOS3; BEM3;* YPL114W*;* YPL113C*; PEX25; CAR1; GDE1;* YPL109C; YPL108W*;* YPL107W*; SSE1*
T37	16L	445842	458799	-0.0902	downstep	*PDR12; SUR1; LCL1; LGE1; LEE1; KTR6*
Radicicol	16R	825645	831055	0.0777	downstep	YPR147C; YPR148C*; NCE102;* YPR150W
Stat	16R	832061	837909	-0.0941	upstep	*URN1;* YPR153W*; PIN3; NCA2*
Stat	16R	882983	883828	-0.1784	upstep	YPR170C; YPR170W-A; YPR170W-B
DMSO	16R	934034	938148	-0.1211	upstep	*SGE1*



Supplementary Figure 3.1: Contribution of individual genes to fitness effects of ROH4 Tamp in nutrient limitation. Fitness as calculated by the change in relative frequency over time of strains with a 220,016 bp amplification on Chromosome 10L and single deletions in (A) phosphate-limitation, (B) Glucose-limitation, and (C) Sulfate-limitation. Regions shaded represent genes present across the breakpoint identified in phosphate-limitation (green) and glucose-limitation (red). Error bars are +/- one standard deviation between replicate chemostats.



Supplementary Figure 3.2: Contribution of individual genes to fitness effects of ROH6 Tamp in nutrient limitation. Fitness as calculated by the change in relative frequency over time of strains with a 136,522 bp amplification on Chromosome 10L and single deletions in (A) phosphate-limitation, (B) Glucose-limitation, and (C) Sulfate-limitation. Regions shaded represent genes present across the breakpoint identified in phosphate-limitation (green). The glucose-limitation breakpoint is not shown, as the Tamp assayed here does not encompass that region. Error bars are +/- one standard deviation between replicate chemostats.

Supplementary Table 3.1: Summary of genes present across breakpoint in phosphate limitation. Shaded region indicates genes within identified breakpoint.

Systematic	Name	Haploid mating	Rescue by mating	Detected in final pool	notes
YJL172W	CPS1	+		+	
YJL171C	TOH1	+		+	
YJL170C	ASG7	+		+	
YJL169W	dubious	+		+	
YJL168C	SET2	+		+	
YJL167W	ERG20	-	+	+	
YJL166W	QCR8	+		+	
YJL165C	HAL5	+		+	
YJL164C	TPK1	-		+	Remade strain
YJL163C	uncharacterized	-		-	unable to confirm barcode
YJL162C	JJJ2	+		+	
YJL161W	FMP33	+		+	
YJL160C	YJL160C	+		+	
YJL159W	HSP150	+		+	
YJL158C	CIS3	+		+	
YJL157C	FAR1	+		+	
YJL156C	SSY5	+	+	+	
YJL155C	FBP26	+		+	
YJL154C	VPS35	+		+	
YJL153C	INO1	+		+	
YJL152W	dubious	+		-	
YJL151C	SNA3	+		+	
YJL150W	dubious	+		+	

Supplementary Table 3.2: Summary of genes present across breakpoint in glucose limitation. Shaded region indicates genes within identified breakpoint.

Systematic	Name	Haploid mating	Rescue by mating	Detected in final pool
YJL149W	DAS1	+		+
YJL148W	RPA34	+		+
YJL147C	MRX5	+		+
YJL146W	IDS2	+		-
YJL145W	SFH5	+		+
YJL144W	ROQ1	+		+
YJL143W	TIM17	+		+
YJL142C	IRC9	+		+
YJL141C	YAK1	+		+
YJL140W	RPB4	+		+
YJL139C	YUR1	+		+
YJL138C	TIF2	+		+
YJL137C	GLG2	-	+	+
YJL136C	RPS21B	+		+
YJL135W	dubious	+		+
YJL134W	LCB3	+		+
YJL133W	MRS3	+		-
YJL132W	putative	+		+
YJL131C	AIM23	-	+	-
YJL130C	URA2	+		+
YJL129C	TRK1	-	+	-
YJL128C	PBS2	+		+
YJL127C	SPT10	+		+
YJL126W	NIT2	+		+
YJL125C	GCD14	+		-
YJL124C	LSM1	+		+
YJL123C	MTC1	+		+
YJL122W	ALB1	+		+
YJL121C	RPE1	+		+
YJL120W	dubious	+		+
YJL119C	dubious	+		+
YJL118W	verified	+		+
YJL117W	PHO86	+		+

Supplementary Table 3.3: Primers used in this study.

Primer ID	Primer Name	Primer Sequence	Notes
OAK-033	YJL086C ROH F1	gatccg ggatcc TGTGTGGTGTGTGGGTGTGTGTGGGTGTGTGGGT GTGTGGG tacgta CTAGGCATCGCTTCTTCGTATG	Amplify region of homology for CFV1 - forward primer
OAK-034	YJL086C ROH R1	catacgccgtagc gacgtc ttgctattcaTGCCAGCACTTTAAGGGTACA	Amplify region of homology for CFV1 - reverse primer
OAK-038	YJL107C ROH F2	gatccg ggatcc TGTGTGGTGTGTGGGTGTGTGTGGGTGTGTGGGT GTGTGGG tacgta TTTCGCCATCCGACGTACAA	Amplify region of homology for CFV4 - forward primer
OAK-039	YJL107C ROH R2	catacgccgtagc gacgtc ggacaagatgAAGAAGATAGGCCTGCTCGC	Amplify region of homology for CFV4 - reverse primer
OAK-041	YJL152W ROH R1	catacgccgtagc gacgtc tgcgcccgccGTAGACCGTGACCTGGAAGC	Amplify region of homology for CFV6 - forward primer
OAK-042	YJL152W ROH F2	gatccg ggatcc TGTGTGGTGTGTGGGTGTGTGTGGGTGTGTGGGT GTGTGGG tacgta TTCGATTCCGCATCCAACCC	Amplify region of homology for CFV6 - reverse primer
OAK-052	barseq forward	ATGGATGTCCACGAGGTCTCT	Truncated PCR barcode for checking deletion collection barcodes
OAK-053	barseq reverse	CAATTCAACGCGTCTGTGAGGGGAGCG	Truncated PCR barcode for checking deletion collection barcodes
OAK-049	barseq seq1	GCGAGGAGCCGTAATTTTTGC	internal sequencing primer for deletion collection barcodes (uptag)
AK8	UP BAR F#	CAAGCAGAAGACGGCATACGAGAT NNNNNN GACCTTACGCGTGGATGTCCACGAGGTCTCT	Amplify deletion collection Uptag barcodes - forward primer
AK9	UP BAR R1	AATGATACGGCGACCACCGAGATCTACAC AGAGGCCGTCACGTCGACCTGCAGCGTACG	Amplify deletion collection Uptag barcodes - reverse primer
AK10	UP SEQ F1	AGA GGC CGT CAC GTC GAC CTG CAG CGT ACG	Illumina sequencing of Uptag barcodes from deletion collection pool - read1
AK11	UP SEQ index	AGAGACCTCGTGGACATCC ACGCGTAAGGTC	Illumina sequencing of Uptag barcodes from deletion collection pool - index1

Supplementary Table 3.4: Strains used in this study.

Strain ID	Strain name	Strain background	Genotype
YAK-099	BY4743	BY4743 (S288C)	MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0
YAK 126	YJL086C_Tamp	BY4743 (S288C)	*MATa/α ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0/LYS+met15Δ0/MET15+can1Δ::LEU2+-MFA1pr-HIS3/CAN1+ YJL086C::kanMX/YJL086C*
YAK 134	YJL107C_Tamp	BY4743 (S288C)	*MATa/α ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0/LYS+met15Δ0/MET15+can1Δ::LEU2+-MFA1pr-HIS3/CAN1+ YJL107C::kanMX/YJL107C*
YAK 142	YJL152W_Tamp	BY4743 (S288C)	*MATa/α ura3Δ0 leu2Δ0 his3Δ1 lys2Δ0/LYS+met15Δ0/MET15+can1Δ::LEU2+-MFA1pr-HIS3/CAN1+ YJL152W::kanMX/YJL152W*
ROH1 pool	Tamp pool 1	BY4743	Heterozygous yeast knockout strains (Supp. Table 3.1 and 3.2) mated to YJL086C_Tamp
ROH4 pool	Tamp pool 2	BY4743	Heterozygous yeast knockout strains (Supp. Table 3.1 and 3.2) mated to YJL107C_Tamp
ROH6 pool	Tamp pool 3	BY4743	Heterozygous yeast knockout strains (Supp. Table 3.1 and 3.2) mated to YJL152WC_Tamp
YMD3004	CFV backbone	*E. coli*	Plasmid: pIL05 (pUC19 backbone) + KlURA3 (ie. *K. lactis* URA3)

Supplementary Table 4.1: Strains used in this study. Base genotype of all strains is S228c *MATa ura3 his3 trp1 leu2 LYS2*

Strain	Alias	Base Ploidy	Chromosome Gains	Chromosome Losses
RLY2626	N	1		
RLY4737	2N	2		
RLY4927	A2	1	II, XII	
RLY4931	A6	2		I, IV, VI, VIII, X, XIII
RLY4932	A7	2		I, III, VI, IX, XIV
RLY4946	A18	2	V, VII, IX, X, XI, XVI	
RLY4948	A20	2	VII, IX, XIII, XIV, XV	
RLY4949	A21	2	II, IV, VII, XIII, XIV, XV	
RLY4950	A22	3		I, III, V, VI, IX, XII, XIII
RLY4951	A23	2	I, II, IX, XI, XII, XIII, XIV, XV	
RLY4952	A24	2	II, VII, IX, XIV	
RLY4962	A34	2	I, IV, VIII, IX, X, XI	
RLY4965	A37	2	II, VII, XIV, XV, XVI	
RLY4966	A38	2	II, IV, VIII, IX, XI, XV	

VITA

Abigail Keller grew up in Woodbury, MN and received a B.S. in Biochemistry from the University of Wisconsin – Madison. While at Wisconsin, she conducted research in the lab of Yevgenya Grinblat investigating the role of *zic2a* and *zic2b* in zebrafish brain and craniofacial development. After graduating, she remained as a full-time Research Intern with the Grinblat Lab, focusing on the role *rfx4* in zebrafish brain development and ciliogenesis. Additionally, Abigail designed and generated novel mutant zebrafish lines using both TALENs and the newly popular CRISPR/Cas9 methods of gene deletion for future use in the Grinblat Lab. She joined the Molecular and Cellular Biology Program at the University of Washington in 2014 and completed her thesis work with Maitreya Dunham in the Genome Sciences Department. Abigail has presented her work at multiple national and international conferences, and has been committed to science outreach from her days with the Biocore Outreach Ambassadors at UW-Madison to her volunteer work at the Seattle Aquarium throughout graduate school.