Assessing Ecology and Evolution of Aneuploidy in Yeast

Eduardo Scopel Ferreira da Costa
October 2019

Major Professor: Douda Bensasson

Committee Member: David Hall
Committee Member: Zachary Lewis
Committee Member: Liang Liu

Important information for the Committee: (from the Graduate Student Handbook)

Instructions for written exam evaluation:
Each committee member has two weeks to review and submit a grade of 'pass' or 'fail' via email to the student and the Graduate Program Administrator (iobgradadmin@UGA.EDU). To pass the written portion and go on to the oral portion, the student must receive no more than one dissenting (failing grade) vote. If a committee member does not provide a grade two weeks after submission of the written exam, the grade will be marked as a 'pass' for that committee member.
The written exam takes the form of an NIH grant proposal; if you are not familiar with the IOB written exam requirements, see Appendix B in the most recent graduate student handbook (https://iob.uga.edu/graduate-program/graduate-handbooks/).

Instructions for oral exam:
The oral exam will last at least two hours, but not longer than three hours. The student will prepare a presentation of no more than 20 slides that are intended to serve as a framework of the discussion of the proposed research. The student's presentation should last for approximately 20-25 minutes without interruptions, followed by questions from the advisory committee and other faculty present. Questions during the exam will test both general and specific knowledge related to the student's proposed research as described in their presentation and written proposal. **A member of the student's committee, other than the advisor, will serve as chair of the exam. The advisor is not allowed to answer questions for the student, and will not participate in the discussion unless granted permission by the exam chair.**

SPECIFIC AIMS

This project's **objective** is to provide a better understanding of the evolutionary role of aneuploidy in yeast, identifying key environmental and genetic factors contributing to its occurrence.

Aneuploidy has major impacts on eukaryotic cells and is usually detrimental to fitness - it is linked to genetic disorders in humans, and slower growth rate in fungi. Nonetheless, aneuploidy is common in eukaryotes (e.g. fungi, protists, animals) and can be advantageous under specific conditions. Many studies suggest aneuploidy is a mechanism of rapid adaptation used by fungi to survive extreme selective pressures. Some examples of phenotypes linked to aneuploidy in fungi are drug-resistance, increased pathogenicity, copper, ethanol, and heat tolerance. The yeast *Saccharomyces cerevisiae* is a great model system to study aneuploidy because it is commonly found as aneuploid; can live in different environments (e.g. beer, humans, trees); is an opportunistic pathogen, making it a good model to study drug-resistance; is one of the most well studied eukaryotes, with many publicly available genome sequences from all over the world. Moreover, despite aneuploidy being common in *S. cerevisiae*, its number of chromosomes has not change in 100 million years, suggesting ploidy state is evolutionary constrained in this species. Recent studies show the frequency of aneuploidy changes among yeast populations, being mostly common in human-associated lineages. Currently, there is a debate over whether aneuploidy is rare or common in wild environments. Resolving this debate is crucial to understand the evolutionary role of aneuploidy in yeast, because wild strains are not under artificial selection, and have been living for a long time in their environments, which could reflect the ancestral habitat of this species.

My **long-term goal** is to develop a model that explains the occurrence of aneuploidy in yeast. In this project my **central hypothesis** is that aneuploidy is rare in wild yeast, and it is only beneficial for local rapid adaptation, being maladaptive in the long-term. My **expectation** is that by the end of this project I will identify key environmental and genetic factors that correlate with an increased frequency of aneuploidy in yeast, building a framework to predict its occurrence in this organism.

Aim 1: Evaluate computational tools to detect aneuploidy from yeast genome data. Here I am going to identify the most appropriate computational approaches to estimate aneuploidy in yeast. Using yeast genome data of verified ploidy, I will assess frequently used approaches to detect ploidy and copy number variation: Read Depth (RD) and B-Allele Frequency (BAF). I expect to provide a baseline for development and improvement of bioinformatic tools to detect aneuploidy.

Aim 2: Identify environmental or genetic factors associated with aneuploidy in *S. cerevisiae*. I hypothesize the frequency of aneuploidy in yeast correlates with genetic background and environment. To test that, I will detect ploidy changes in over 2,000 yeast strains isolated from different environments all around the world. I expect aneuploidy to be rare in wild and common in human-associated environments, and find common environmental and genetic patterns linked to a high frequency of aneuploidy.

Aim 3: Test fitness effects of aneuploidy in different environments. One example of aneuploidy as a transient response to extreme directional selective pressures in yeast is drug-resistance in *Candida albicans*. When exposed to anti-fungal drugs, *C. albicans* gain an extra copy of the chromosome harboring resistance genes. The aneuploid cells are eventually replaced in the population by euploids over-expressing genes giving the desired phenotype. I hypothesize that aneuploidy is beneficial in the short-term, but maladaptive in the long-term. To test that, I will experimentally evolve strains from different lineages under conditions representing human-associated and wild environments. I expect euploids to be more generally fit than aneuploids.

This project will help me understand the evolutionary role of aneuploidy in yeast, and identify the main factors behind its occurrence, providing a baseline for development of a predictive model.

RESEARCH STRATEGY

A. SIGNIFICANCE

Aneuploidy is an uneven number of chromosomes in a cell that happens via gain or loss of one or more chromosomes during cell division. It is an important genetic variation linked to many diseases in humans, and pathogenic phenotypes in microbes. For example, aneuploidy is linked to cancer (metastasis, drug resistance, disease progression, and tumorigenesis) and genetic disorders (Down's Syndrome) in humans, and drug-resistance phenotypes in fungal pathogens, such as *C. albicans* and *Cryptococcus neoformans* (Dunlap et al., 1986; Selmecki et al., 2006; Torres et al., 2007; Pavelka et al., 2010; Sionov et al., 2010; Kwon-Chung and Chang, 2012; Gerstein and Berman, 2015; Berman, 2016; Gao et al., 2016; Sansregret and Swanton, 2017; Tsai and Nelliat, 2019; Chunduri and Storchova, 2019). In the plant pathogens *Fusarium oxysporum* and *Nectria haematococca*, aneuploidy seems to be a requirement for pathogenicity (Miao et al., 1991; Ma et al., 2010), and some parasites, like *Giardia*, *Leishmania*, and Cythrids are frequently aneuploids (Lachaud et al., 2014; Refsnider et al., 2015; Tumova et al., 2016).

Aneuploidy also impacts the baker's yeast, *S. cerevisiae*, which is considered an opportunistic pathogen (Pérez-Torrado and Querol, 2016). *S. cerevisiae* live in many different environments (Peter et al., 2018), and can survive as aneuploids in lab, industrial, clinical, and wild environments (Ezov et al., 2006; Yona et al., 2012; Tan et al., 2013; Hose et al., 2015; Gallone et al., 2016; Zhu et al., 2016; Peter et al., 2018; Duan et al., 2018; Fay et al., 2019; Morard et al., 2019). Despite this common presence in eukaryotes, aneuploidy usually decreases fitness in all organisms where it is present (Torres et al., 2008; Santaguida and Amon, 2015; Gilchrist and Stelkens, 2019). Additionally, in the lab, aneuploidy sometimes arises in yeast cells when stressful conditions are imposed (Selmecki et al., 2009; Yona et al., 2012; Chen et al., 2012; Selmecki et al., 2015; Wertheimer et al., 2016), but aneuploid cells tend to be replaced by euploids once those conditions are gone (Tang and Amon, 2013; Hose et al., 2015; Berman, 2016; Gilchrist and Stelkens, 2019). These findings lead to the generalization that aneuploidy is a key adaptation mechanism to extreme directional selective pressures in fungi (Todd et al., 2017).

Different environmental conditions probably have a significant influence on how yeast respond to stress. Yeast populations living in wild environments, such as woodlands, are under nutrient-poor substrates, fluctuating temperatures and humidity (diurnal and seasonal), and competing with other species, conditions that can trigger their sexual cycle through sporulation (meiosis) (Liti, 2015). In most human-associated environments conditions are less harsh. Beer and sake strains, for example, grow under sterile continuous cultures, with plenty of food and in stable, controlled physical-chemical conditions. They still have to survive some stressful conditions, due to intra-specific competition and sudden shutdowns, but they do not need to sporulate (Bokulich and Bamforth, 2013). In contrast, wine strains sporulate more often (Marsit and Dequin, 2015), probably due to the presence of harsher selective pressures present in winemaking. Wine fermentation happens at low pH, yields high levels of ethanol and sulfur dioxide, and permits the presence of other microbial species that thrive in the beginning of the process, when ethanol concentrations are low and sugar is abundant (Fugelsang and Edwards, 2007). Not surprisingly, strains isolated from oak trees sporulate better than wine and industrial strains (Gerke et al., 2006; Marsit and Dequin, 2015; Kong et al., 2018), whereas some sake and beer strains do not sporulate at all (Anderson and Martin, 1975; Fukuda et al., 2016). One thing wine and wild environments have in common is seasonality, while most other human-associated environments are non-seasonal. Interestingly, my preliminary data (Figure 1) show aneuploidy is rare in seasonal and common in non-seasonal sterile environments that permit continuous growth of yeast. This is indicative of a relationship between high frequency of aneuploidy in environments where sporulation is not com-

mon (beer and sake) - spores generated from aneuploid cells are often inviable (Deutschbauer et al., 2002).

By comparing the frequency of aneuploidy in populations isolated from different environments, it is possible to identify environmental conditions associated with an increased occurrence of aneuploidy. My preliminary analysis shows that the (i) frequency of aneuploidy in yeast significantly changes among environments, thus suggesting environmental factors correlate with this mutation (Figure 1); (ii) seasonality and brewing conditions are important environmental factors that can explain the occurrence of aneuploidy in diploid yeast. If aneuploidy is just a mechanism fungi use to cope with short-term directional selective pressures (e.g. in beer, sake), it could be maladaptive in seasonal environments where harsher and more diverse stressful conditions are present (e.g. trees, wine), explaining the discrepancy among these environments.

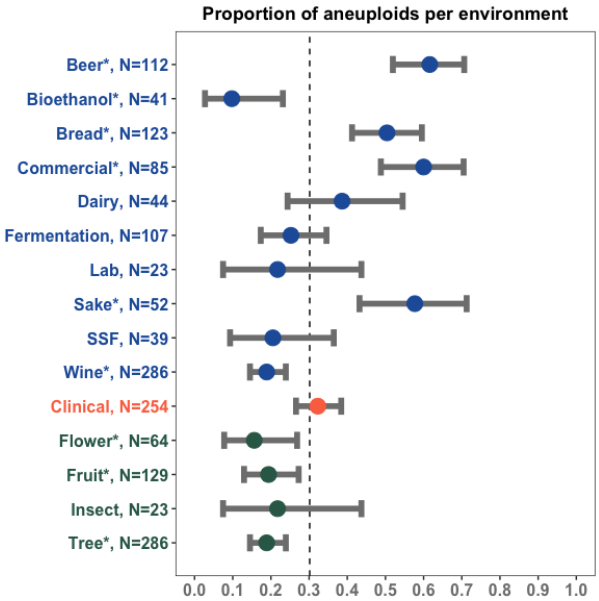

Figure 1: Aneuploidy is rare in seasonal (Wild, Wine and Bioethanol) and common in human-associated, non-seasonal environments (Beer, Bread, Commercial, Sake). Environments shown in blue are human-associated (non-clinical); orange are human-associated (clinical); and green are wild. Dots and bars show proportion of aneuploids in each environment, with 95% confidence intervals from the binomial distribution. Dotted line shows the average proportion of aneuploids.

The main objectives of this project are to improve computational approaches to detect ploidy changes in fungi, study the evolutionary role of aneuploidy in yeast, and identify key environmental and genetic factors behind its occurrence. Upon completion, *this project will resolve the debate of whether aneuploidy is rare or common in wild environments, and identify key explanatory factors correlated to its appearance in yeast.* **This contribution will be significant because it will shine light on the long-term evolutionary relevance of aneuploidy**, which would be indicative of the evolutionary state of a particular strain. It will also bring me closer to the development of a predictive model that can estimate the probability of aneuploidy to occur in yeast given a specific genetic background and environmental conditions.

B. INNOVATION

A deep analysis of aneuploidy in yeast focusing on environments has not been done so far, to the extent of my knowledge. Leveraging the many yeast genome sequences publicly available, I will study the occurrence of aneuploidy in strains of *S. cerevisiae* isolated from diverse environments, with different evolutionary histories, and genetic backgrounds. The innovative aspect of this project is that it will investigate **the occurrence of aneuploidy in yeast under an evolutionary perspective, considering phylogenetics, environmental and genetic factors.** The evolutionary role of aneuploidy in environmental yeast has not been clarified to date. On a different note, I am going to **use experimental evolution to test how aneuploidy affects fitness of yeast strains from different lineages under environmental conditions associated with high and low levels of aneuploidy.** Even though experimental evolution assays have been done on lab yeast strains (Dunham et al., 2002; Gresham et al., 2008; Yona et al., 2012; Lu et al., 2016; Scott et al., 2017; Tsai et al., 2019), none focused on how fitness and ploidy of specific genetic backgrounds respond to wild-like and human-associated environments. This combination of experimental evolution and genome-wide association studies to identify genetic and environmental predictors of aneuploidy

Aneuploidy						
Tools	VarScan2[1]	ControlFreec[1]	nQuire[2]	ploidyNGS[2]	vcf2alleleplot[2]	Ymap[1,2]
S. cerevisiae (N = 23 genomes)						
Sensitivity	84%	94%	76%	60%	64%	82%
Specificity	100%	98%	69%	99%	100%	99%
FDR	0%	23%	86%	12%	0%	13%
C. albicans (N = 23 genomes)						
Sensitivity	86%	86%	82%	92%	86%	86%
Specificity	100%	98%	92%	99%	100%	98%
FDR	0%	25%	55%	14%	0%	20%

Table 1: Overall performance of RD and BAF tools. [1] - Read Depth; [2] - B-Allele Frequency. $Sensitivity = (TruePositives)/(TruePositives + FalsePositives)$; $Specificity = (TrueNegatives)/(TrueNegatives + FalsePositives)$; $FDR(FalseDiscoveryRate) = (FalsePositives)/(TruePositives + FalsePositives)$

is novel. Not only this will help me understand how specific strains become aneuploid in the first place, but I will also learn its consequences in short and long-term adaptation.

C. APPROACH

C.1. Preliminary Studies

Motivation for Specific Aim 1. In preparation for Specific Aim 2, I needed to test how well computation approaches, Read Depth (RD), and B-Allele Frequency (BAF), to detect aneuploidy would work on genome data from *S. cerevisiae*.

RD detects aneuploidy comparing the average number of reads that align to the reference genome between chromosomes. If the number of reads is uniformly distributed across the genome the organism is euploid, but if there is a significant deviation in one or more chromosomes, it is aneuploid. BAF counts the number of SNPs that differ from the reference genome (B-alleles - or alternate alleles), and calculates its ratio from the total number of alleles in that position. From this ratio, heterozygous sites can be found in the genome, indirectly estimating copy number. As I aimed at detecting aneuploidy on over 2,000 strains with varied genomic structures, it was essential to understand any technical limitations and biases these methods had. To do that, I designed Specific Aim 1, that focused on evaluating computational methods to detect aneuploidy in yeast. This Aim is almost complete.

RD and BAF are complementary approaches to detect aneuploidy. My preliminary results show that: 1) RD is versatile, but dependent on sequencing methods and quality. This method works well on yeast species, regardless of their ploidy level and heterozygosity, and can detect segmental copy number variations. However, it fails to detect some aneuploidies when the sequencing coverage is low or somehow biased (e.g. by DNA fragmentation method). Therefore, I do not see differences between yeast species *S. cerevisiae* and *C. albicans* when evaluating RD tools (Table 1). 2) BAF can detect overall ploidy, but it is contingent on the level of heterozygosity. Detection of overall ploidy is a major advantage of BAF as RD cannot do that. Unfortunately, BAF can only estimate overall ploidy when the strain has enough heterozygous sites in the genome. Another strong feature of BAF is it works well on poor-quality sequences. As heterozygosity is crucial for this approach to work, BAF's accuracy vary a lot between species (Table 1). When RD and BAF are combined the accuracy of aneuploidy detection is over 90%.

BAF can detect contamination in genome data. In collaboration with Brent Shuman from the Momany Lab, we found known homozygous strains of *S. cerevisiae* and *Aspergillus fumigatus* to have many heterozygous sites when analyzed with BAF, with allele ratios consistent to what would be expected for decaploids (BAF = 0.9). The fact that decaploidy is very unlikely to appear in known homozygous haploid and diploid strains made us hypothesize that the genome data was contaminated (a 10% contamination would give a BAF of 0.9). Therefore, we decided to

test if BAF could be used to detect contamination in genome sequencing data, and what would be the consequences of such contamination in downstream analyses, such as detecting ploidy changes, population genomics, and phylogenetics. Our preliminary findings are that even a 10% contamination rate would reduce the availability of high-quality SNPs by almost 70%, regardless of the strain's ploidy. It also makes it more difficult to detect aneuploidy in contaminated heterozygous strains using BAF.

Frequency of aneuploidy changes among environments in *S. cerevisiae*. To determine the key environmental factors behind the occurrence of aneuploidy in yeast, I investigated over 2,000 yeast genome sequences, and counted the number of aneuploids and euploids in each of 19 environmental categories - classified according to the substrates strains were isolated from (Figure 5). I found the frequency of aneuploidy to be especially high in some domesticated, non-seasonal environments, such as beer, bread, commercial and sake; and low in seasonal environments, such as wine, and wild (Figure 1 above).

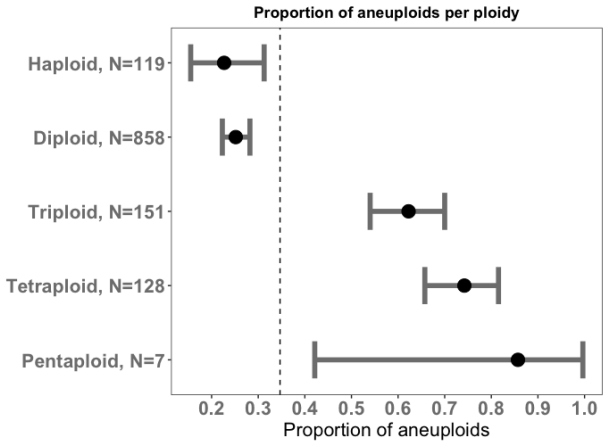

Figure 2: Aneuploidy is common in polyploids. Dots represent proportion of aneuploidy per ploidy; bars represent 95% confidence intervals from the binomial distribution; dotted line represents average proportion of aneuploidy.

Aneuploidy associates with polyploidy and heterozygosity in *S. cerevisiae*. Analyzing my data I confirmed that aneuploidy is more frequent in polyploid strains (Figure 2), in agreement with other studies (Gilchrist and Stelkens, 2019). Polyploidy is more frequent in some environments, such as beer and bread (Selmecki et al., 2015; Fay et al., 2019), what could explain why aneuploidy is more common in these strains. This could happen because aneuploidy can be an intermediate state between polyploidy and euploidy (Berman, 2016), and polyploid strains frequently generate aneuploid progeny (Potapova et al., 2013). Considering this association between polyploidy and aneuploidy, I tested the association between environments and aneuploidy on natural diploid strains only. Moreover, as suggested by (Peter et al., 2018), my findings confirm that heterozygosity correlates with frequency of aneuploidy, regardless of genetic background and environment. The number of aneuploids increases especially when heterozygosity is greater than 0.3% (Figure 3), the average pairwise distance between strains from different continents (Liti et al., 2009). Some of the more heterozygous strains in my data set are possibly hybrids between *S. cerevisaie* and a closely related *Saccharomyces* species, such as *S. paradoxus*, also common on trees (Robinson et al., 2016). Besides, ploidy instability is a known mechanism of fertility recovery in yeast hybrids (Gilchrist and Stelkens, 2019; Charron et al., 2019), which could also explain this correlation between heterozygosity and aneuploidy.

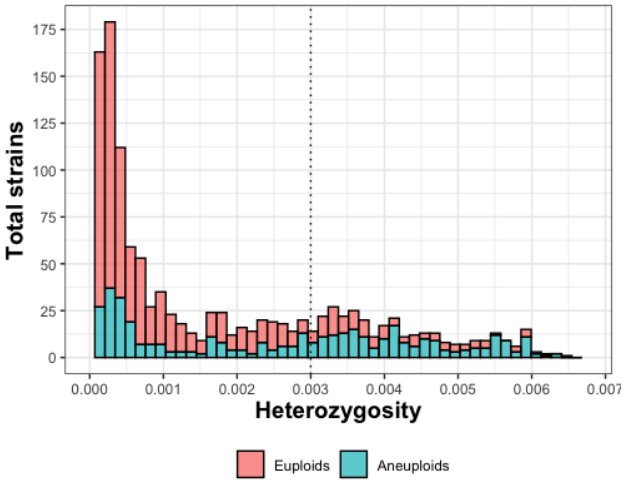

Figure 3: Highly heterozygous strains are more frequently aneuploid. Vertical dotted line represents the average pairwise distance between strains from different continents (Liti et al., 2009). Strains with heterozygosity greater than 1% were excluded.

Seasonality seems to be a key environmental factor. To investigate the association between

aneuploidy and environment, I used a logistic regression (generalized linear model - glm) with aneuploidy (binary) as response variable, heterozygosity (continuous) and environment (categorical) as explanatory variables on natural diploid yeast strains, with a binomial error distribution. The minimal adequate model suggests that the frequency of aneuploidy increases in Brewing (Beer and Sake) and decreases in Seasonal (Wild and Wine) and other 7 (Bioethanol, Bread, Commercial, Clinical, Dairy, Fermentation, SSF) environments, that are too heterogeneous to be grouped (Figure 4). One caveat of this result is that the variance among variables in this model is not homogeneous (Fligner-Killeen test, df = 8, p-value = 1.249e-06), most likely due to the number of highly heterozygous strains in seasonal environments (Figure 4). Most of the highly heterozygous beer strains are polyploids - excluded here. In order to better predict the association between aneuploidy and environment I may exclude possible hybrids from my data set.

C.2. Research Plan

Specific Aim 1: Evaluate computational tools to detect aneuploidy from yeast genome data.

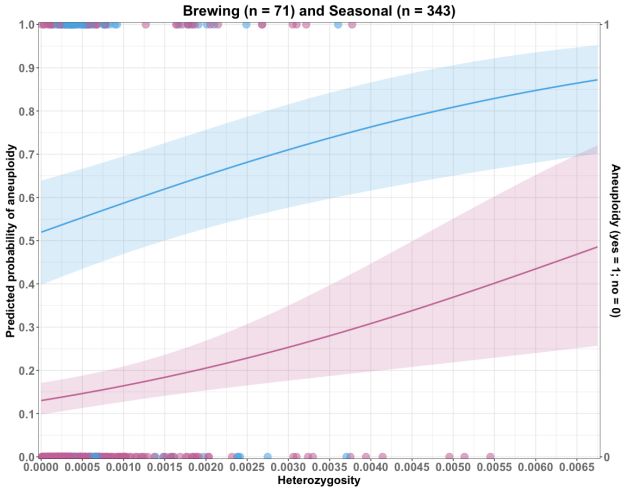

Figure 4: Logistic regression predicts aneuploidy increases in Brewing (Blue) and decreases in Seasonal (Pink) environments. Lines represent predicted probabilities. Shaded areas represent confidence intervals. Dots show aneuploidy (yes=1; no=0). The minimal model includes 7 other environments not shown here.

Knowing the ploidy state of an organism is essential to study its genomics, because any ploidy changes interfere with gene expression. Several methods can be used to estimate ploidy: Karyotyping, Flow Cytometry, qPCR, Pulsed-Field Gel Electrophoresis (PFGE), Read Depth (RD), and B-Allele Frequency (BAF) (Gorter de Vries et al., 2017). While the first 4 are wet lab methods, RD and BAF are computational approaches that use genome sequencing data only. With the wide availability of genome data and low sequencing costs, it makes sense to consider computational methods in lieu of more time consuming and costly wet lab experiments. While there are plenty of established tools to measure copy number variation, most of them focus on cancer (Koboldt et al., 2012; Liu et al., 2013; Pabinger et al., 2014; Nam et al., 2015; Zare et al., 2017), fetal cells (Fan et al., 2008; Sehnert et al., 2011; Lo et al., 2014), segmental copy number variation (Koike et al., 2011; Teo et al., 2012; Duan et al., 2013; Zhao et al., 2013; Guo et al., 2013; Liu et al., 2013; Samarakoon et al., 2014; Tan et al., 2014; Pirooznia et al., 2015; Li et al., 2018) and exome data (Sathirapongsasuti et al., 2011; Love et al., 2011; Plagnol et al., 2012; Koboldt et al., 2012; Bao et al., 2014; Wang et al., 2014; Kong et al., 2017; Tan et al., 2017). Most of these tools were extensively benchmarked in the literature. However, one that focus on eukaryotes with diverse genome structures, such as plants, fungi, protists, and ancient organisms (with low quality DNA) is lacking. Furthermore, those tools require prior knowledge of overall ploidy to work well, which is a safe assumption for humans, but not necessarily for other organisms.

The **objective** of this aim was to assess the efficiency of computational tools to estimate aneuploidy from yeast genome data. The tools I selected here are based on the most commonly used approaches: Read Depth (RD), and B-Allele Frequency (BAF). This aim did not have a specific working hypothesis, as it was supposed to benchmark a set of tools on diverse yeast genome sequences. My **expectation** was to highlight the main advantages and challenges of each tool, focusing on the specific caveats to be considered when evaluating ploidy variation in fungi. My **approach** was to test six RD and BAF tools (Table 2) on a very diverse data set of 26 strains

of two yeast species: *S. cerevisiae* and *C. albicans*. The strains selected have a wide range of ploidy states (from haploids to tetraploids), levels of heterozygosity (homozygous to highly heterozygous), and are from different sequencing projects, using different DNA library preparations. The **rationale** for this aim was to identify the best tools to use, and what problems to expect when detecting aneuploidy on my data set for Specific Aim 2. Also, it serves as a guideline for the development of new tools to detect changes in ploidy focusing on fungal organisms.

Selecting RD and BAF tools. I assessed the most frequent computational approaches to estimate ploidy variation from genome sequencing data (Table 2), using popularity (most cited tools), versatility (ability to detect multiple levels of ploidy), application (used on more than one species), and recentness (just for BAF, as there were not many highly cited tools) as my selection criteria. Most of the other BAF tools described in the literature are not versatile enough (do not consider ploidy changes above triploidy), or rely on other methods to detect ploidy variation.

Tools	Method	Citations	Application	Reference
VarScan2	Read Depth	2,310	Cancer	(Koboldt et al., 2012)
Control-Freec	Both	386	Cancer	(Boeva et al., 2011)
nQuire	B-Allele Frequency	14	Fungi	(Weiß et al., 2018)
ploidyNGS	B-Allele Frequency	16	Fungi	(Augusto Corrêa dos Santos et al., 2017)
vcf2alleleplot	B-Allele Frequency	14	Yeast	(Bensasson et al., 2019)
Ymap	Both	30	Yeast	(Abbey et al., 2014)

Table 2: Description of ploidy/CNV calling tools evaluated in this study.

Building a varied data set with yeast. I used a total of 46 strains: 23 from *C. albicans* and 23 from *S. cerevisiae*, with different ploidy states and levels of heterozygosity, including euploids and aneuploids, and from different sequencing projects (Strope et al., 2015; Hirakawa et al., 2015; Peter et al., 2018; Ropars et al., 2018). The idea here was to test the tools described in Table 2 with a wide range of strains, to push their limits to the maximum. The strains used here had their ploidy level previously confirmed by a combination of flow cytometry and computational tools.

Evaluating accuracy, sensitivity, and specificity. To test each tool described in Table 2 with my data set, I measured their ability to correctly call ploidy changes by calculating their *sensitivity* (ratio between the number of True Positives and Total Positives, including False Positives), *specificity* (ratio between the number of True Negatives and Total Negatives, including False Negatives), and *false discovery rate* (ratio between the number of False Positives and the sum of True and False Positives), as shown in Table 1. By doing that, I identified the strengths and weaknesses of each program, providing examples of successes and failures of each one. Finally, I discussed their use on organisms with different genome structures, highlighting when their use is more or less recommended. Differently from other approaches, I focused not only on the characteristics of the tools, but also on how they differ when applied to organisms with different genome structure (see Preliminary Studies).

Expected outcomes. Here I defined that the most appropriate approach to detect ploidy variation on yeast is to combine RD and BAF for heterozygous, and RD only for homozygous strains. I showed that BAF accuracy varies with heterozygosity (Table 1), failing to estimate ploidy on homozygous strains. I found that the genome sequencing method interferes with RD, mainly due to biases introduced by library preparation methods. Finally, I showed that BAF detects contamination in genome data, which is being further investigated in a collaboration with Brent Shuman from the Momany Lab.

Potential problems and alternative strategies. Dealing with homozygous strains is a challenge that is only partially dealt by RD. Even though detection of overall ploidy has improved in the past few years, the accuracy of computational tools is not great. Except for homozygous strains, the best solution is to use both methods together. By doing this, the accuracy goes over 90%. It might be possible to detect overall ploidy of homozygous *S. cerevisiae* strains by looking at the allele ratio in their mating locus only. If the strain is haploid the allele ratio in that region should be

around 1.0, whereas diploid strains will have a BAF of 0.5. This is a future direction I want to take in this project.

Specific Aim 2: Identify environmental or genetic factors associated with aneuploidy in *S. cerevisiae*.

Even though aneuploidy is common in *S. cerevisiae* (Peter et al., 2018), the number of chromosomes in this species has not changed in the past 100 million years (Wolfe and Shields, 1997). The fact that aneuploidy is usually found in human-associated environments, such as clinical, beer, and sake (Gallone et al., 2016; Zhu et al., 2016; Peter et al., 2018; Fay et al., 2019), and may be rare in wild environments (Peter et al., 2018; Duan et al., 2018), raises the question if aneuploidy is only useful for rapid local adaptation, being maladaptive in the long-term. To answer this question, it is crucial to know if aneuploidy is rare (Peter et al., 2018) or common (Duan et al., 2018) in wild environments. My preliminary results (Figure 1) suggest that not only aneuploidy is rare in wild, but also in wine and bioethanol production, all seasonal environments. In contrast, aneuploidy is common in non-seasonal (beer and sake) environments, where continuous growth is permitted (Figures 1 and 4). Seasonality induces sporulation (meiosis) in yeast (Liti, 2015), so it is not surprising that wild and wine strains sporulate much better than beer and sake strains (Anderson and Martin, 1975; Gerke et al., 2006; Bokulich and Bamforth, 2013; Fukuda et al., 2016; Kong et al., 2018). Furthermore, when aneuploid cells try to sporulate the result is almost always deleterious (Deutschbauer et al., 2002). Thus, I propose to investigate the relationship of seasonal and non-seasonal environments with the frequency of aneuploidy in *S. cerevisiae*. The **objective** of this aim is to identify environmental or genetic factors associated with an increased frequency of aneuploidy in *S. cerevisiae*.

Categorizing strains into environmental categories. After identifying over 2,000 *S. cerevisiae* genome sequences publicly available (Skelly et al., 2013; Bergström et al., 2014; Almeida et al., 2015; Song et al., 2015; Strope et al., 2015; Barbosa et al., 2016; Borneman et al., 2016; Gallone et al., 2016; Gayevskiy et al., 2016; Gonçalves et al., 2016; Zhu et al., 2016; Yue et al., 2017; Barbosa et al., 2018; Duan et al., 2018; Peter et al., 2018; Fay et al., 2019), I categorized them into 19 environmental categories, based on the substrates they were isolated from, according to metadata available in the sequencing projects (Figure 5). Strains isolated from unknown or non-specific substrates were excluded from the analysis. Also, strains that are used as starter cultures for beer, bread, sake or wine production were categorized as *Commercial*.

Environment	# of strains
Wine	363
Commercial (including starter cultures)	223
Beer	184
Liquid Fermentation	136
Bread	129
Sake	59
Dairy	47
Solid-state fermentation (SSF)	45
Bioethanol	43
Lab	28
Coffee	22
Cocoa	18
Industrial	10
Clinical	277
Trees	192
Fruit	156
Flower	74
Insect	26
Other plants	16
Total	**2048**

Figure 5: 2,048 strains isolated from known substrates categorized into 19 environments. Blue represents non-clinical human-associated, orange clinical, and green wild environments.

Estimating ploidy variation and heterozygosity. I detected ploidy variation (polyploidy, partial, and whole-chromosome aneuploidy) in the strains described in the previous section using a combination of RD and BAF. First, I mapped each strain's genome sequence to the *S. cerevisiae* reference genome (S288c), counted the number of reads that aligned to the reference (for RD), and detected variant sites (for BAF). Then, I used in-house scripts (RD) and vcf2alleleplot (BAF) (Bensasson et al., 2019) to detect ploidy changes. Heterozygosity was measured with vcf2alleleplot. By the end

of this analysis, each strain had a numerical value for overall ploidy (1 to 7), a binary value for aneuploidy (yes/no), a continuous value for heterozygosity, a categorical value for environment (19 factors - Figure 5), a numerical value [1..16] for gain/loss of specific chromosome(s), and a categorical value for genetic lineage (based on phylogenetic estimates from other studies). This step prepared the data set for the association analysis (glm). The results from this analysis are shown in Figures 1, 2, and 3.

Determining if aneuploidy associates with environmental or genetic factors. The main goal here is to identify and quantify any correlation between the frequency of aneuploidy with environment or genetic background. My preliminary results suggest an association between polyploidy, heterozygosity and aneuploidy (Figures 2 and 3), so including those factors as explanatory variables would be problematic. To deal with this issue, I excluded from the data set non-diploids and known hybrids (heterozygosity greater than 1%). Then I used a logistic regression (glm) with aneuploidy (binary) as a response variable, and heterozygosity and environment as explanatory variables. I simplified the full model (with 14 ecological categories, and an interaction term between environment and heterozygosity), following the methods described in Crawley (2012), until I got a minimal adequate model without the heterozygosity:environment interaction term, and with 9 environmental factors - Beer and Sake were grouped into the Brewing category; and Trees, Fruit, Flower, Insect, and Wine were grouped into the Seasonal category. Before accepting this as the minimal model, I tested if other simplifications would explain the model better. First, I checked if Competition was more relevant than Seasonality as an environmental factor by grouping Bioethanol and/or Dairy with the Seasonal group, reducing to 7 environmental factors. To test if Continuous Growth was more important than Brewing, I grouped Bioethanol with the Brewing group, reducing to 8 environmental factors - Beer, Bioethanol, and Sake are all produced in continuous cultures. In all cases the model with the Seasonal group (9 categories) was significantly better. These results suggest an association between Seasonality and Brewing with low and high frequencies of aneuploidy, respectively. However, at least one other additional factor must be included in the analysis: Genetic Lineage. Moreover, from my preliminary tests, the variance between variables seems to heterogeneous, which would make glm not the best approach. This needs a more careful investigation though, especially considering the possible presence of hybrids among the strains. If glm is not adequate after excluding highly admixed strains and including Genetic Lineage as an explanatory variable, I can use Linear Mixed Models (Price et al., 2010; Earle et al., 2016), considering Genetic Lineage as a random effect, or Deep Learning (Zou et al., 2019; Eraslan et al., 2019), leveraging the extensive SNPs data set I have available. I expect to confirm my preliminary findings.

Identifying which chromosomes are more frequently gained or lost. Chromosomes that are more likely to gain or lose copies might harbor genes that when differentially expressed increase fitness. In this step, I will count aneuploidies per chromosome. Next, I will test if there is an association between the frequency of aneuploidy per chromosome and environment using a generalized linear model (Crawley, 2012) - aneuploidy (binary) as response variable, chromosome (16 factors), and environment (14 factors) as explanatory variables. By associating the gain/loss of specific chromosomes with environments, I will be able to investigate differentially expressed genes (by copy number variation) in these aneuploid chromosomes. Using the *Saccharomyces* Genome Database (SGD) (Cherry et al., 2011), I will identify phenotypes linked to these genes and infer potential fitness advantages aneuploid strains living in the associated environments may have. This experiment will give me candidate genes that might be targets for selection in human-associated and wild environments. I expect to find that some chromosomes are more likely to become aneuploids, and others to be more restrained.

Investigating meiosis genes in wild and beer strains. If sporulation is not common in non-seasonal environments, genes associated with sporulation may have lost function in strains adapted to these environments. I will use SGD to search for sporulation genes and build a phylogenetic tree based on them, with wild and beer strains. I will determine how different sporulations genes are between strains isolated from seasonal and non-seasonal environments. My expectation is that these genes are conserved in wild strains, but not in beer ones. Also, I will annotate the best phylogenetic tree available with each aneuploidy event found in this aim. Upon completion, I will recognize when aneuploidy happened in each lineage. My expectation is that aneuploidy is a more transient event, appearing only occasionally in the tree.

Expected outcomes. I expect the frequency of aneuploidy to change among environments, more specifically, aneuploidy will be rare in wild environments and common in human-associated ones. There seems to be an association between seasonality, brewing and aneuploidy, that I expect to confirm. Furthermore, I expect to find genetic factors associated with habitats where aneuploidy is common. For example, in Brewing environments, chromosomes harboring sporulation genes might be more frequently aneuploids than in Seasonal environments. These outcomes will help me understand if there is a selective constraint maintaining euploidy in wild yeast, and what is the long-term evolutionary role of aneuploidy in yeast.

Potential problems and alternative strategies. One complicating factor of this experiment is the association of aneuploidy with polyploidy and heterozygosity (Figures 2 and 3). My plan to overcome the polyploidy dependence is to split my experiments by ploidy level. It makes sense to analyze different ploidy levels separately because aneuploidy arises in diploids and polyploids by different biological reasons. Aneuploidy is common in polyploid strains because it is an intermediate step to diploidy (Berman, 2016). The strategy to resolve the heterozygosity dependence is similar. Highly heterozygous strains could be inter-specific hybrids. This is a problem because hybridization in yeast commonly results in aneuploidy due to chromosome mis-segregation in meiosis caused by lack of chromosomal crossover (Boynton et al., 2018), or due to a recently found ploidy instability in hybrids (Charron et al., 2019). As these events are only expected in hybrids or highly admixed strains it is reasonable to treat them differently. To that end I can adjust the heterozygosity filter (e.g. from 1% to 0.3%) on my data set and analyze these strains separately.

Specific Aim 3: Test fitness effects of aneuploidy in different environments.

Aneuploidy is a common mutation in yeast. It is usually detrimental to fitness under replete conditions but can be adaptive when specific stressful conditions are present (Bennett et al., 2014; Todd et al., 2017), and it is common for aneuploidy to be lost when selective pressures are no longer present (Bennett et al., 2014; Berman, 2016). This transience of aneuploidy suggests it may not be beneficial in the long-term in yeast. My **objective** in this aim is to test how beer and wild lineages of aneuploid and euploid *S. cerevisiae* strains respond, in terms of ploidy changes and fitness, to (i) temperature fluctuations; and (ii) nutrient depletion. I chose these general selective pressures based on my preliminary findings suggesting seasonality reduces and brewing increases the frequency of aneuploidy in yeast (Figures 1 and 4). My **working hypothesis** is euploid cells from wild lineages grow faster on seasonal and nutrient-depleted (wild-like) environments, and at least as fast as aneuploid cells in replete, stable conditions (brewing-like). Furthermore, I expect aneuploidy to be lost in the long-term, regardless of environmental conditions. My **approach** to test these hypotheses will be to measure how euploid and aneuploid strains from Beer, Wild, and Lab lineages respond, in terms of fitness, to wild-like environments (Aim 3.a); compare how euploid and aneuploid Wild strains evolve through time under selection pressures that are representative of brewing and wild-like environments (Figure 6), measuring changes in ploidy and fitness (Aim 3.b); identify the most common mutations (loss or gain of aneuploidy, loss of heterozygosity, and

SNPs) arising in these environments (Aim 3.c).

Strain selection from beer and wild lineages. Phenotypic changes are affected by strain genetic background (Sardi et al., 2018), and the same is true for ploidy change (Todd et al., 2017). To account for that, in my experiments I will use *S. cerevisiae* strains from distinct Wild and Beer lineages. The most well represented wild strains were isolated from oak trees, and they cluster under several lineages, such as Mediterranean Oak, North Carolina Oak, China, and Mosaic (Almeida et al., 2015; Peter et al., 2018; Duan et al., 2018). In aims 3.a and 3.b I will use 8 wild strains isolated from oak trees: 2 aneuploids and 2 euploids from Mosaic and 2 aneuploids and 2 euploids from non-Mosaic (North America and Europe) lineages. Inclusion of mosaic lineages allows me to detect if admixture has any effects in phenotypic and ploidy responses to environmental conditions. Beer strains are mainly split into two phylogenetically distinct groups: Beer 1 and Beer 2 (Gallone et al., 2016), therefore it makes sense to use 1 aneuploid and 1 euploid strain of each one of these lineages for experiments in Aim 3.a.

3.a. Measuring phenotypic differences between strains living in wild and beer-like environments. I will compare how 9 yeast strains (aneuploid and euploid representatives of Beer 1, Beer 2, North America Oak, and European Oak lineages, plus a Lab control, such as S288c) respond to wild (solid nutrient-poor media) and brewing-like (liquid nutrient-rich media) environments. More specifically, I will inoculate 10^6-10^7 cells of each one of these 9 strains separately on 3 different conditions, and grow them for 72 hours: (1) Oak Bark Infusion Agar solid media (OBIA) (Belotte et al., 2003; Kowallik et al., 2015) at 30℃; (2) Yeast Peptone Dextrose-Agar solid media (YPDA) at 30℃; (3) Yeast Peptone Dextrose liquid media (YPD), at 30 ℃. Conditions (1) and (3) represent wild and brewing-like environments respectively, and (2) is a control for (1). For conditions (1) and (2) I will first grow the strains on liquid YPD in a 96-well plate for 24 hours, then transfer then to the solid media (YPDA or OBIA) in a petri dish, using a 48 pin multi-blot replicator. Each strain will have 3 replicates, so each petri dish will have 27 colonies. I will replicates each petri dish 3 times, to account for OBIA media variability, so in the end there will 81 colonies separated in 3 petri dishes for each one of conditions (1) and (2). Strains growing on condition (3) will be grown in 96-well plates. At 24, 48, and 72 hours I will estimate growth rate by comparing colony sizes, using Quantitative Fitness Analysis Workflow (QFA), as described by Banks et al. (2012) for conditions (1) and (2), and by measuring Optical Density using a microplate reader for condition (3). I will also measure sporulation efficiency at 0 hours, by dissecting at least 60 tetrads of each strain (Börner and Cha, 2015). By the end of this experiment, I expect wild euploid strains to grow faster in the wild-like environment (condition 1), and at least as fast as the beer strains in conditions (2) and (3). I expect euploid strains to grow faster than aneuploids under all conditions.

3.b. Determining how selective pressures associated with wild and brewing environments affect wild strains of *S. cerevisiae*. I will use experimental evolution to compare how 8 aneuploid and euploid wild strains, plus 1 lab control, respond, in terms of fitness and ploidy, to brewing-like (constant temperature, and nutrient-rich media) and wild-like (fluctuant temperature, nutrient-poor media) environments. More specifically, I will grow each one of the 9 strains (1 aneuploid and 1 euploid from non-mosaic North America Oak; 1 aneuploid and 1 euploid from non-mosaic European Oak; 2 aneuploids and 2 euploids from Mosaic lineage; 1 lab strain, such as S288c) individually under the following conditions: (A) YPDA, at 30℃; (B) YPDA, with daily temperature shifts (from 30℃ to 15℃); (C) modified minimal synthetic media with 0.5% glucose, instead of 2.0% (SD-agar), at 30℃; (D) SD-agar, with daily temperature shifts (from 30℃ to 15℃). I will transfer grown colonies to fresh media every 24 hours using replica plating technique (Lederberg and Lederberg, 1952), until they reach 1,000 generations (Figure 6). I chose this number because aneuploidy arose frequently (75%) in *S. cerevisiae* strains growing under glucose-limiting chemostat in less

than 500 generations, according to Dunham et al. (2002). As chemostats yield population sizes greater than ones in solid media colonies, I doubled the number of generations.

Determining the duplication time of each strain. Yeast strains with different genetic backgrounds and ploidy states have different phenotypes (Thorburn et al., 2013; Sardi et al., 2018), which includes their doubling time. That is especially true if they are growing under different environmental conditions (Slater et al., 1977). Therefore, I expect the strains I am using in this aim to have different doubling times when growing on various conditions. To correctly estimate the number of daily transfers in the experimental evolution assay, I will determine the doubling time of each strain growing on environments A, B, C and D. To do that I will inoculate 10^6-10^7 cells/mL of each strain into environments A, B, C and D, and grow them for 72 hours, separately. Every 24 hours, I will count the number of cells using a hemocytometer chamber or by comparing colony sizes with QFA, and calculate the time it takes for each strain to double the number of cells. By doing that, I will know the number of generations each strain produces daily, which is essential to decide the number of daily transfers in the experimental evolution assay.

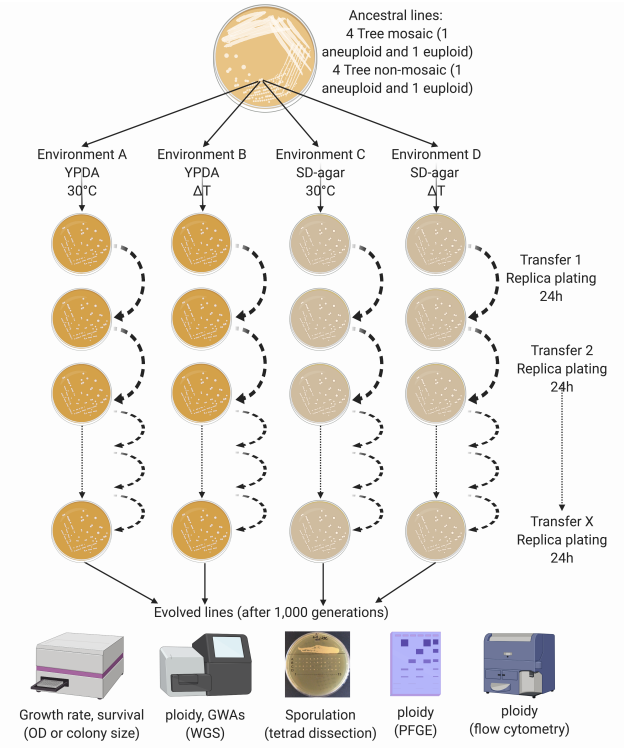

Figure 6: Tree strains will be inoculated into 4 different environments and experimentally evolved for 1,000 generations. I will measure the evolved lines phenotypes (growth rate and sporulation), ploidy (flow cytometry, and PFGE) and mutations (via WGS). Number of transfers will be calculated in Aim 3.b.

Testing detection efficiency of gain/loss of aneuploidy. As a pilot test, I will experimentally evolve each of the 8 wild strains for at least 100 generations under two conditions: (i) solid YPDA, at 30℃, (ii) solid YPDA, at 30℃, with Fluconazole $32\mu g/ml$. After 100 generations I will detect ploidy changes in these strains with PFGE, and if there are none I will continue the experiment for 100 more generations, until I find at least 10% of aneuploid colonies. My expectation is to determine how often gain/loss of aneuploidy happens in control conditions so I can adapt the number of generations in my main experiment, if necessary. If gain of aneuploidy is still too rare in the presence of fluconazole, I will pretreat cells with $20\mu g/ml$ of radiciclol, which should increased the number of aneuploids dramatically (Chen et al., 2012). I expect loss of aneuploidy to be common in condition (i) and gain to be common in condition (ii).

Detecting ploidy changes under different environmental conditions. I expect aneuploidy to be lost in most environments where aneuploid strains are growing, due to the appearance of new euploid cells that have a fitness advantage over them. It is also possible for aneuploidy to arise in environments where temperature shifts are common, due to the sudden stress the cells are going through. To detect any ploidy changes, at every 100 generations I will karyotype (by PFGE and Flow Cytometry) 1 or 2 replicates of each strain growing under environments A, B, C, and D.

Comparing fitness between wild strains living in wild and beer-like environments. Similarly to Aim 3.a, I will investigate how individual strains respond to the selective pressures imposed by each environment. At the end of 1,000 generations I will measure each strain's sporulation efficiency (by dissecting at least 60 tetrads). At every 100 generations I will measure their growth rate (by

cell counting, and QFA) and freeze one or two colonies to make stock cultures of different time-points. Lines that evolved to sporulate better, or developed some other interesting phenotype, will be further investigated with the analysis of frozen stocks. I expect euploid strains to grow faster, and sporulate better than aneuploids in all environments.

3.c. Genome-wide analysis of experimentally evolved lines. I will generate genome sequences for 1 or 2 replicates of each experimentally evolved line after 1,000 generations, for no more than 72 sequenced genomes. The genomes will be sequenced to 50x coverage at the Georgia Genomics and Bioinformatics Core (GGBC), and I will use the pipelines described here (Aims 1 and 2) and in Bensasson et al. (2019) (adapted to *S. cerevisiae*) to detect ploidy changes, loss of heterozygosity, SNPs, and other mutations. I will associate these mutations with phenotypes and ploidy changes arising in the experimentally evolved lines, and detect changes between lineages, environments and initial ploidy states. Additionally, I will use phenotypic and genotypic variation obtained in the previous steps to build a model explaining the association between genotypes and aneuploidy, like described in Aim 2 (Crawley, 2012).

Expected outcomes. This project will identify how selective pressures associated with wild-like environments affect ploidy and fitness in wild and beer yeast. The beer strains are probably very well adapted to the brewing environment, but will not grow very well in deplete conditions. Euploid cells will probably have an advantage in harsher environments, and should be at least as fit as aneuploids under replete conditions. For the wild lineages my expectation is that euploids will grow faster and sporulate better than aneuploid strains in all environments, in the long-term. I expect some aneuploid cells to eventually arise in environments where suddens shifts in temperature are common, but they should be replaced by euploid cells in the long-term. By detecting ploidy changes every few generations and in the end of the experimental evolution assay, I should be able to detect these transient gains of aneuploidy, and subsequent losses in the long-term.

Potential problems and alternative strategies. It is possible for beer strains to not grow at all in OBIA media, and that I cannot collect enough oak bark to make enough media for all the proposed experiments in aim 3.a. To overcome that problem, I can replace OBIA media by SD-agar, used in Aim 3.b, without losing the poor-nutrient characteristic of environments (1) and (2). Another potential problem I can find, especially in environments where temperature shifts are common, is bottleneck effects. To avoid this, I will measure the average growth rate of each genetic background on each condition to adapt the time between transfers (aim 3.b) - if a particular strain is growing slower on a specific condition it might not generate enough cells to surpass the bottleneck population size (10^6-10^7 cells) (Dunham, 2010).

Timeline

Goals	Year 1 (17-18)	Year 2 (18-19)	Year 3 (19-20)	Year 4 (20-21)	Year 5 (21-22)
Aim 1: Evaluate computational tools					
Selecting RD and BAF tools and build yeast data set					
Evaluating accuracy, sensitivity, and specificity of tools					
Aim 2: Identify environmental or genetic factors					
Categorizing strains into environments and estimating ploidy, heterozygosity					
Associating aneuploidy with environments					
Investigating meiosis genes					
Aim 3: Test fitness effects of aneuploidy in different environments					
Measuring fitness and ploidy in beer strains under different environments					
Experimental evolution of wild strains, measuring ploidy and fitness					
Genome-wide analysis of experimentally evolved lines					

References

Abbey, D. A., Funt, J., Lurie-Weinberger, M. N., Thompson, D. A., Regev, A., Myers, C. L., and Berman, J. (2014). Y map: a pipeline for visualization of copy number variation and loss of heterozygosity in eukaryotic pathogens. *Genome medicine*, 6(11):100.

Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J.-L., Serra, M., Dequin, S., Couloux, A., et al. (2015). A population genomics insight into the mediterranean origins of wine yeast domestication. *Molecular ecology*, 24(21):5412–5427.

Anderson, E. and Martin, P. A. (1975). The sporulation and mating of brewing yeast. *J. Inst. Brew.*, 81:242–247.

Augusto Corrêa dos Santos, R., Goldman, G. H., and Riaño-Pachón, D. M. (2017). ploidyngs: visually exploring ploidy with next generation sequencing data. *Bioinformatics*, 33(16):2575–2576.

Banks, A., Lawless, C., and Lydall, D. (2012). A quantitative fitness analysis workflow. *JoVE (Journal of Visualized Experiments)*, (66):e4018.

Bao, L., Pu, M., and Messer, K. (2014). Abscn-seq: a statistical method to estimate tumor purity, ploidy and absolute copy numbers from next-generation sequencing data. *Bioinformatics*, 30(8):1056–1063.

Barbosa, R., Almeida, P., Safar, S. V., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., et al. (2016). Evidence of natural hybridization in brazilian wild lineages of saccharomyces cerevisiae. *Genome biology and evolution*, 8(2):317–329.

Barbosa, R., Pontes, A., Santos, R. O., Montandon, G. G., de Ponzzes-Gomes, C. M., Morais, P. B., Gonçalves, P., Rosa, C. A., and Sampaio, J. P. (2018). Multiple rounds of artificial selection promote microbe secondary domestication?the case of cachaca yeasts. *Genome biology and evolution*, 10(8):1939–1955.

Belotte, D., Curien, J. B., Maclean, R. C., and Bell, G. (2003). An experimental test of local adaptation in soil bacteria. *Evolution*, 57(1):27–36.

Bennett, R. J., Forche, A., and Berman, J. (2014). Rapid mechanisms for generating genome diversity: Whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harbor Perspectives in Medicine*, 4(10).

Bensasson, D., Dicks, J., Ludwig, J. M., Bond, C. J., Elliston, A., Roberts, I. N., and James, S. A. (2019). Diverse lineages of candida albicans live on old oaks. *Genetics*, 211(1):277–288.

Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., et al. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular biology and evolution*, 31(4):872–888.

Berman, J. (2016). Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS yeast research*, 16(3):fow020.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., and Barillot, E. (2011). Control-freec: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, 28(3):423–425.

Bokulich, N. A. and Bamforth, C. W. (2013). The microbiology of malting and brewing. *Microbiol. Mol. Biol. Rev.*, 77(2):157–172.

Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., and Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of saccharomyces cerevisiae. *G3: Genes, Genomes, Genetics*, 6(4):957–971.

Börner, G. V. and Cha, R. S. (2015). Analysis of yeast sporulation efficiency, spore viability, and meiotic recombination on solid medium. *Cold Spring Harbor Protocols*, 2015(11):pdb–prot085027.

Boynton, P. J., Janzen, T., and Greig, D. (2018). Modeling the contributions of chromosome segregation errors and aneuploidy to saccharomyces hybrid sterility. *Yeast*, 35(1):85–98.

Charron, G., Marsit, S., Henault, M., Martin, H., and Landry, C. (2019). Spontaneous whole-genome duplication restores fertility in interspecific hybrids. *bioRxiv*, page 538298.

Chen, G., Bradford, W. D., Seidel, C. W., and Li, R. (2012). Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature*, 482(7384):246.

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., et al. (2011). Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic acids research*, 40(D1):D700–D705.

Chunduri, N. K. and Storchova, Z. (2019). The diverse consequences of aneuploidy. *Nature cell biology*, 21(1):54–62.

Crawley, M. J. (2012). *The R Book*. Wiley Publishing, 2nd edition.

Deutschbauer, A. M., Williams, R. M., Chu, A. M., and Davis, R. W. (2002). Parallel phenotypic analysis of sporulation and postgermination growth in saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 99(24):15530–15535.

Duan, J., Zhang, J.-G., Deng, H.-W., and Wang, Y.-P. (2013). Comparative studies of copy number variation detection methods for next-generation sequencing technologies. *PloS one*, 8(3):e59128.

Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., and Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from far east asia. *Nature communications*, 9(1):2690.

Dunham, M. J. (2010). Experimental evolution in yeast: a practical guide. In *Methods in enzymology*, volume 470, pages 487–507. Elsevier.

Dunham, M. J., Badrane, H., Ferea, T., Adams, J., Brown, P. O., Rosenzweig, F., and Botstein, D. (2002). Characteristic genome rearrangements in experimental evolution of saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 99(25):16144–16149.

Dunlap, S. S., Aziz, M., and Rosenbaum, K. (1986). Comparative anatomical analysis of human trisomies 13, 18, and 21: I. the forelimb. *Teratology*, 33(2):159–186.

Earle, S. G., Wu, C.-H., Charlesworth, J., Stoesser, N., Gordon, N. C., Walker, T. M., Spencer, C. C., Iqbal, Z., Clifton, D. A., Hopkins, K. L., et al. (2016). Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nature microbiology*, 1(5):16041.

Eraslan, G., Avsec, Ž., Gagneur, J., and Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, page 1.

Ezov, T. K., Boger-Nadjar, E., Frenkel, Z., Katsperovski, I., Kemeny, S., Nevo, E., Korol, A., and Kashi, Y. (2006). Molecular-genetic biodiversity in a natural population of the yeast saccharomyces cerevisiae from "evolution canyon": microsatellite polymorphism, ploidy and controversial sexual status. *Genetics*, 174(3):1455–1468.

Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L., and Quake, S. R. (2008). Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing dna from maternal blood. *Proceedings of the National Academy of Sciences*, 105(42):16266–16271.

Fay, J. C., Liu, P., Ong, G. T., Dunham, M. J., Cromie, G. A., Jeffery, E. W., Ludlow, C. L., and Dudley, A. M. (2019). A polyploid admixed origin of beer yeasts derived from european and asian wine populations. *PLoS biology*, 17(3):e3000147.

Fugelsang, K. and Edwards, C. (2007). *Wine Microbiology: Practical Applications and Procedures*. Springer US.

Fukuda, N., Kaishima, M., Ishii, J., Kondo, A., and Honda, S. (2016). Continuous crossbreeding of sake yeasts using growth selection systems for a-type and α-type cells. *AMB Express*, 6(1):45.

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., et al. (2016). Domestication and divergence of saccharomyces cerevisiae beer yeasts. *Cell*, 166(6):1397–1410.

Gao, C., Su, Y., Koeman, J., Haak, E., Dykema, K., Essenberg, C., Hudson, E., Petillo, D., Khoo, S. K., and Woude, G. F. V. (2016). Chromosome instability drives phenotypic switching to metastasis. *Proceedings of the National Academy of Sciences*, 113(51):14793–14798.

Gayevskiy, V., Lee, S., and Goddard, M. R. (2016). European derived saccharomyces cerevisiae colonisation of new zealand vineyards aided by humans. *FEMS yeast research*, 16(7):fow091.

Gerke, J. P., Chen, C. T. L., and Cohen, B. A. (2006). Natural isolates of saccharomyces cerevisiae display complex genetic variation in sporulation efficiency. *Genetics*, 174(2):985–997.

Gerstein, A. C. and Berman, J. (2015). Shift and adapt: the costs and benefits of karyotype variations. *Current opinion in Microbiology*, 26:130–136.

Gilchrist, C. and Stelkens, R. (2019). Aneuploidy in yeast: Segregation error or adaptation mechanism? *Yeast*.

Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., and Sampaio, J. P. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Current Biology*, 26(20):2750–2761.

Gorter de Vries, A. R., Pronk, J. T., and Daran, J.-M. G. (2017). Industrial relevance of chromosomal copy number variation in saccharomyces yeasts. *Applied and Environmental Microbiology*, 83(11).

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

Guo, Y., Sheng, Q., Samuels, D. C., Lehmann, B., Bauer, J. A., Pietenpol, J., and Shyr, Y. (2013). Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control. *BioMed research international*, 2013.

Hirakawa, M. P., Martinez, D. A., Sakthikumar, S., Anderson, M. Z., Berlin, A., Gujja, S., Zeng, Q., Zisson, E., Wang, J. M., Greenberg, J. M., et al. (2015). Genetic and phenotypic intra-species variation in candida albicans. *Genome research*, 25(3):413–425.

Hose, J., Yong, C. M., Sardi, M., Wang, Z., Newton, M. A., and Gasch, A. P. (2015). Dosage compensation can buffer copy-number variation in wild yeast. *Elife*, 4:e05462.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., and Wilson, R. K. (2012). Varscan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*, 22(3):568–576.

Koike, A., Nishida, N., Yamashita, D., and Tokunaga, K. (2011). Comparative analysis of copy number variation detection methods and database construction. *BMC genetics*, 12(1):29.

Kong, I. I., Turner, T. L., Kim, H., Kim, S. R., and Jin, Y.-S. (2018). Phenotypic evaluation and characterization of 21 industrial saccharomyces cerevisiae yeast strains. *FEMS yeast research*, 18(1):foy001.

Kong, J., Shin, J., Won, J., Lee, K., Lee, U., and Yoon, J. (2017). Excnvss: A noise-robust method for copy number variation detection in whole exome sequencing data. *BioMed research international*, 2017.

Kowallik, V., Miller, E., and Greig, D. (2015). The interaction of saccharomyces paradoxus with its natural competitors on oak bark. *Molecular Ecology*, 24(7):1596–1610.

Kwon-Chung, K. J. and Chang, Y. C. (2012). Aneuploidy and drug resistance in pathogenic fungi. *PLoS pathogens*, 8(11):e1003022.

Lachaud, L., Bourgeois, N., Kuk, N., Morelle, C., Crobu, L., Merlin, G., Bastien, P., Pagès, M., and Sterkers, Y. (2014). Constitutive mosaic aneuploidy is a unique genetic feature widespread in the leishmania genus. *Microbes and infection*, 16(1):61–66.

Lederberg, J. and Lederberg, E. M. (1952). Replica plating and indirect selection of bacterial mutants. *Journal of bacteriology*, 63(3):399.

Li, S., Dou, X., Gao, R., Ge, X., Qian, M., and Wan, L. (2018). A remark on copy number variation detection methods. *PloS one*, 13(4):e0196226.

Liti, G. (2015). The natural history of model organisms: The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *eLife*, 4:e05835.

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., et al. (2009). Population genomics of domestic and wild yeasts. *Nature*, 458(7236):337.

Liu, B., Morrison, C. D., Johnson, C. S., Trump, D. L., Qin, M., Conroy, J. C., Wang, J., and Liu, S. (2013). Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. *Oncotarget*, 4(11):1868.

Lo, K. K., Boustred, C., Chitty, L. S., and Plagnol, V. (2014). Rapidr: an analysis package for non-invasive prenatal testing of aneuploidy. *Bioinformatics*, 30(20):2965–2967.

Love, M. I., Myšičková, A., Sun, R., Kalscheuer, V., Vingron, M., and Haas, S. A. (2011). Modeling read counts for cnv detection in exome sequencing data. *Statistical applications in genetics and molecular biology*, 10(1).

Lu, Y.-J., Swamy, K. B., and Leu, J.-Y. (2016). Experimental evolution reveals interplay between sch9 and polyploid stability in yeast. *PLoS genetics*, 12(11):e1006409.

Ma, L.-J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in fusarium. *Nature*, 464(7287):367.

Marsit, S. and Dequin, S. (2015). Diversity and adaptive evolution of saccharomyces wine yeast: a review. *FEMS yeast research*, 15(7):fov067.

Miao, V. P., Covert, S. F., and VanEtten, H. D. (1991). A fungal gene for antibiotic resistance on a dispensable (" b") chromosome. *Science*, 254(5039):1773–1776.

Morard, M., Macías, L. G., Adam, A. C., Lairón-Peris, M., Pérez-Torrado, R., Toft, C., and Barrio, E. (2019). Aneuploidy and ethanol tolerance in saccharomyces cerevisiae. *Frontiers in genetics*, 10:82.

Nam, J.-Y., Kim, N. K., Kim, S. C., Joung, J.-G., Xi, R., Lee, S., Park, P. J., and Park, W.-Y. (2015). Evaluation of somatic copy number estimation tools for whole-exome sequencing data. *Briefings in bioinformatics*, 17(2):185–192.

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M. R., Zschocke, J., and Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in bioinformatics*, 15(2):256–278.

Pavelka, N., Rancati, G., and Li, R. (2010). Dr jekyll and mr hyde: role of aneuploidy in cellular adaptation and cancer. *Current opinion in cell biology*, 22(6):809–815.

Pérez-Torrado, R. and Querol, A. (2016). Opportunistic strains of saccharomyces cerevisiae: a potential risk sold in food products. *Frontiers in microbiology*, 6:1522.

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 saccharomyces cerevisiae isolates. *Nature*, 556(7701):339.

Pirooznia, M., Goes, F. S., and Zandi, P. P. (2015). Whole-genome cnv analysis: advances in computational approaches. *Frontiers in genetics*, 6:138.

Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., Grigoriadou, S., Wood, N. W., Hambleton, S., Burns, S. O., Thrasher, A. J., et al. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, 28(21):2747–2754.

Potapova, T. A., Zhu, J., and Li, R. (2013). Aneuploidy and chromosomal instability: a vicious cycle driving cellular evolution and cancer genome chaos. *Cancer and Metastasis Reviews*, 32(3-4):377–389.

Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11(7):459.

Refsnider, J. M., Poorten, T. J., Langhammer, P. F., Burrowes, P. A., and Rosenblum, E. B. (2015). Genomic correlates of virulence attenuation in the deadly amphibian chytrid fungus, batrachochytrium dendrobatidis. *G3: Genes, Genomes, Genetics*, 5(11):2291–2298.

Robinson, H. A., Pinharanda, A., and Bensasson, D. (2016). Summer temperature can predict the distribution of wild yeast populations. *Ecology and Evolution*, 6(4):1236–1250.

Ropars, J., Maufrais, C., Diogo, D., Marcet-Houben, M., Perin, A., Sertour, N., Mosca, K., Permal, E., Laval, G., Bouchier, C., et al. (2018). Gene flow contributes to diversification of the major fungal pathogen candida albicans. *Nature communications*, 9(1):2253.

Samarakoon, P. S., Sorte, H. S., Kristiansen, B. E., Skodje, T., Sheng, Y., Tjønnfjord, G. E., Stadheim, B., Stray-Pedersen, A., Rødningen, O. K., and Lyle, R. (2014). Identification of copy number variants from exome sequence data. *BMC genomics*, 15(1):661.

Sansregret, L. and Swanton, C. (2017). The role of aneuploidy in cancer evolution. *Cold Spring Harbor perspectives in medicine*, 7(1):a028373.

Santaguida, S. and Amon, A. (2015). Short-and long-term effects of chromosome mis-segregation and aneuploidy. *Nature reviews Molecular cell biology*, 16(8):473.

Sardi, M., Paithane, V., Place, M., Robinson, D. E., Hose, J., Wohlbach, D. J., and Gasch, A. P. (2018). Genome-wide association across saccharomyces cerevisiae strains reveals substantial variation in underlying gene requirements for toxin tolerance. *PLOS Genetics*, 14(2):1–25.

Sathirapongsasuti, J. F., Lee, H., Horst, B. A., Brunner, G., Cochran, A. J., Binder, S., Quackenbush, J., and Nelson, S. F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: Exomecnv. *Bioinformatics*, 27(19):2648–2654.

Scott, A. L., Richmond, P. A., Dowell, R. D., and Selmecki, A. M. (2017). The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source. *Molecular biology and evolution*, 34(10):2690–2703.

Sehnert, A. J., Rhees, B., Comstock, D., de Feo, E., Heilek, G., Burke, J., and Rava, R. P. (2011). Optimal detection of fetal chromosomal abnormalities by massively parallel dna sequencing of cell-free fetal dna from maternal blood. *Clinical chemistry*, 57(7):1042–1049.

Selmecki, A., Forche, A., and Berman, J. (2006). Aneuploidy and isochromosome formation in drug-resistant candida albicans. *Science*, 313(5785):367–370.

Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B., and Berman, J. (2009). Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS genetics*, 5(10):e1000705.

Selmecki, A. M., Maruvka, Y. E., Richmond, P. A., Guillet, M., Shoresh, N., Sorenson, A. L., De, S., Kishony, R., Michor, F., Dowell, R., et al. (2015). Polyploidy can drive rapid adaptation in yeast. *Nature*, 519(7543):349.

Sionov, E., Lee, H., Chang, Y. C., and Kwon-Chung, K. J. (2010). Cryptococcus neoformans overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. *PLoS pathogens*, 6(4):e1000848.

Skelly, D. A., Merrihew, G. E., Riffle, M., Connelly, C. F., Kerr, E. O., Johansson, M., Jaschob, D., Graczyk, B., Shulman, N. J., Wakefield, J., et al. (2013). Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome research*, 23(9):1496–1504.

Slater, M. L., Sharrow, S. O., and Gart, J. J. (1977). Cell cycle of saccharomycescerevisiae in populations growing at different rates. *Proceedings of the National Academy of Sciences*, 74(9):3850–3854.

Song, G., Dickins, B. J., Demeter, J., Engel, S., Dunn, B., and Cherry, J. M. (2015). Agape (automated genome analysis pipeline) for pan-genome analysis of saccharomyces cerevisiae. *PLoS One*, 10(3):e0120671.

Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., and McCusker, J. H. (2015). The 100-genomes strains, an s. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome research*, 25(5):762–774.

Tan, R., Wang, J., Wu, X., Juan, L., Zheng, L., Ma, R., Zhan, Q., Wang, T., Jin, S., Jiang, Q., et al. (2017). Erds-exome: a hybrid approach for copy number variant detection from whole-exome sequencing data. *IEEE/ACM transactions on computational biology and bioinformatics*.

Tan, R., Wang, Y., Kleinstein, S. E., Liu, Y., Zhu, X., Guo, H., Jiang, Q., Allen, A. S., and Zhu, M. (2014). An evaluation of copy number variation detection tools from whole-exome sequencing data. *Human mutation*, 35(7):899–907.

Tan, Z., Hays, M., Cromie, G. A., Jeffery, E. W., Scott, A. C., Ahyong, V., Sirr, A., Skupin, A., and Dudley, A. M. (2013). Aneuploidy underlies a multicellular phenotypic switch. *Proceedings of the National Academy of Sciences*, 110(30):12367–12372.

Tang, Y.-C. and Amon, A. (2013). Gene copy-number alterations: a cost-benefit analysis. *Cell*, 152(3):394–405.

Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S., and Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics*, 28(21):2711–2718.

Thorburn, R. R., Gonzalez, C., Brar, G. A., Christen, S., Carlile, T. M., Ingolia, N. T., Sauer, U., Weissman, J. S., and Amon, A. (2013). Aneuploid yeast strains exhibit defects in cell growth and passage through start. *Molecular biology of the cell*, 24(9):1274–1289.

Todd, R. T., Forche, A., and Selmecki, A. (2017). Ploidy variation in fungi–polyploidy, aneuploidy, and genome evolution. *Microbiology spectrum*, 5(4).

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

Torres, E. M., Williams, B. R., and Amon, A. (2008). Aneuploidy: cells losing their balance. *Genetics*, 179(2):737–746.

Tsai, H.-J. and Nelliat, A. (2019). A double-edged sword: Aneuploidy is a prevalent strategy in fungal adaptation. *Genes*, 10(10):787.

Tsai, H.-J., Nelliat, A. R., Choudhury, M. I., Kucharavy, A., Bradford, W. D., Cook, M. E., Kim, J., Mair, D. B., Sun, S. X., Schatz, M. C., et al. (2019). Hypo-osmotic-like stress underlies general cellular defects of aneuploidy. *Nature*, 570(7759):117.

Tumova, P., Uzlikova, M., Jurczyk, T., and Nohynkova, E. (2016). Constitutive aneuploidy and genomic instability in the single-celled eukaryote giardia intestinalis. *MicrobiologyOpen*, 5(4):560–574.

Wang, H., Nettleton, D., and Ying, K. (2014). Copy number variation detection using next generation sequencing read counts. *BMC bioinformatics*, 15(1):109.

Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., and Burbano, H. A. (2018). nquire: a statistical framework for ploidy estimation using next generation sequencing. *BMC bioinformatics*, 19(1):122.

Wertheimer, N. B., Stone, N., and Berman, J. (2016). Ploidy dynamics and evolvability in fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1709):20150461.

Wolfe, K. H. and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387(6634):708.

Yona, A. H., Manor, Y. S., Herbst, R. H., Romano, G. H., Mitchell, A., Kupiec, M., Pilpel, Y., and Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proceedings of the National Academy of Sciences*, 109(51):21010–21015.

Yue, J.-X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., et al. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature genetics*, 49(6):913.

Zare, F., Dow, M., Monteleone, N., Hosny, A., and Nabavi, S. (2017). An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC bioinformatics*, 18(1):286.

Zhao, M., Wang, Q., Wang, Q., Jia, P., and Zhao, Z. (2013). Computational tools for copy number variation (cnv) detection using next-generation sequencing data: features and perspectives. *BMC bioinformatics*, 14(11):S1.

Zhu, Y. O., Sherlock, G., and Petrov, D. A. (2016). Whole genome analysis of 132 clinical saccharomyces cerevisiae strains reveals extensive ploidy variation. *G3: Genes, Genomes, Genetics*, 6(8):2421–2434.

Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., and Telenti, A. (2019). A primer on deep learning in genomics. *Nature genetics*, 51(1):12–18.