Host determinants of 2-micron plasmid stability in *Saccharomyces cerevisiae*

Michelle Jeanne Hays

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Harmit S. Malik, Chair

Aimée M. Dudley

M.K. Raghuraman

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Host determinants of 2-micron stability in *Saccharomyces cerevisiae*

Michelle Jeanne Hays

Chair of the Supervisory Committee:
Harmit Singh Malik, PhD
Division of Basic Sciences, Fred Hutchinson Cancer Research Center

Selfish genetic elements are DNA parasites that exploit their host cells for their own reproduction, thereby reducing host fitness. How can host cells evolve to defend themselves against these genetic parasites? I seek to address this question using the 2-micron plasmid, a selfish element naturally found across budding yeasts. This plasmid hijacks host cellular machinery to replicate and segregate itself, resulting in a 1-3% fitness cost to the host. Despite this cost, most *Saccharomyces cerevisiae* isolates carry the plasmid, indicating that it is a remarkably successful, co-evolved genetic parasite of yeasts. I hypothesized that some *S. cerevisiae* strains may have evolved the ability to restrict the plasmid and thereby evade parasitism. By screening a panel of natural isolates, I identified three strains that naturally do not harbor the 2-micron. I find that

when the plasmid is reintroduced in the laboratory, these strains reproducibly lose the 2-micron, indicating that plasmid loss is a heritable trait. Furthermore, this plasmid loss phenotype is a genetically dominant trait, supporting the hypothesis that these strains have evolved a restriction factor targeting the 2-micron plasmid. I took a QTL mapping strategy to identify a genomic locus underlying plasmid loss. Additionally, I have developed a rapid plasmid loss assay that facilitates monitoring of plasmid occupancy in live cells across a population at single cell resolution. This assay is higher throughput and can better explore population heterogeneity compared to traditional plasmid loss methods. This work will allow us to explore the genetics, molecular mechanism, and possible fitness tradeoffs underlying a naturally evolved parasite resistance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisory committee for all their time, encouragement and support. In particular Dr. Aimée M. Dudley and Dr. M.K. Raghuraman who are my reading committee, and Dr. Bonny Brewer: I have enjoyed all of my time meeting with you and getting to join your lab meetings over the years. I appreciate how generous you have all been with your time and ideas! Thank you to Paula Levan and Cynthis Wong for their help in the laboratory creating strains, performing screens, and providing excellent conversation and a fun time. I would also like to thank Patrick Nugent and Sam Hart for their hard work while rotating. Thank you to the entire Malik lab community for the mentoring and support. You have been essential to the development of this project. Finally, thank you Dr. Harmit Malik for the encouragement, intellectual freedom, supportive environment and fun.

DEDICATION

This work is dedicated to Hank, Olivia and Momo.

Chapter 1. INTRODUCTION

1.1 SELFISH GENETIC ELEMENTS

Host genomes are engaged in longstanding conflicts with a myriad of selfish genetic elements (SGE)[1,2,3]. One class of SGEs, including viruses, plasmids and other pathogens, invade a host cell and coopt host cellular machinery for their own propagation. Host genomes have had to evolve several defense strategies to block or control these invading SGEs. For example, bacterial and archaeal cells deploy restriction endonucleases and CRISPR-Cas systems to defend against phage and other mobile genetic elements[4]. Similarly, mammalian cells encode restriction factors such as TRIM5alpha and APOBEC genes to block the spread of invading viruses[5]. Recent work suggests that our understanding and discovery of such protective defense systems is rudimentary. For example, in *Saccharomyces cerevisiae*, one of the best characterized model eukaryotes, the only defined restriction factor (*XRN1*) was described as recently as 2016[6]. It is surprising that yeast cellular immunity is so poorly characterized, given how well studied yeasts are in general and how little horizontal transmission of SGEs have been observed amongst budding yeasts[7].

A second class of SGEs, including transposons, exist within the host's genome[3]. This integration leads to leads to intragenomic conflict when the selfish element's drive to propagate is at odds with the rest of the genome. As such, molecular mechanisms have evolved to protect the host genome from the over-proliferation of transposable elements[8]. In animals, much of this protection occurs in the germline (because those genomes go on to the next generation), which becomes a battleground for the arm-race between host genomes and their resident transposons[9]. Other SGEs in this class can escape Mendelian inheritance to increase their own transmission to future generations. For example, many SGEs create post-segregation

dysfunction in competitor cells to increase their transmission. In bacteria this can occur via 'toxin-antidote' systems in which host cells become "addicted" to the selfish element, and cells that lose the SGE succumb to the toxic effects of the encoded toxin[10]. In eukaryotes, such post-segregation dysfunction manifests after meiosis, where haploid cells destined to become pollen, sperm, or spores are specifically prevented from doing so via toxic mechanisms encoded on competitor chromosomes. Many examples of this selfish behavior have been discovered, including distinct spore-killers in multiple fungi and Segregation Distorter in Drosophila[11–13]. Even essential chromosomal elements can act as SGEs by manipulating meiosis[14]. For example, centromeres in plants and animals can influence their non-Mendelian transmission by outcompeting homologs for inclusion in the oocyte[15].

1.2 ARMS RACES BETWEEN SGES AND HOST GENOMES

Despite their varied mechanisms, all SGEs increase their own reproductive fitness within a cell or population, at the expense of the host's own fitness[1]. This antagonistic relationship between SGEs and host genomes often shapes the evolutionary trajectory of associated genes. For example, most SGEs rely on host cellular machinery to replicate their genomes, to transcribe and translate their proteins, and to ensure their proliferation by passage into new cells[1,3]. However, this parasitism of host machinery can lower host reproductive fitness. If a host variant arises that can suppress selfish elements, this variant is favored by natural selection, and can spread in a population over time. In turn, variants of genetic parasites able to evade this restriction will be selected for, allowing for continued survival within a host[16].

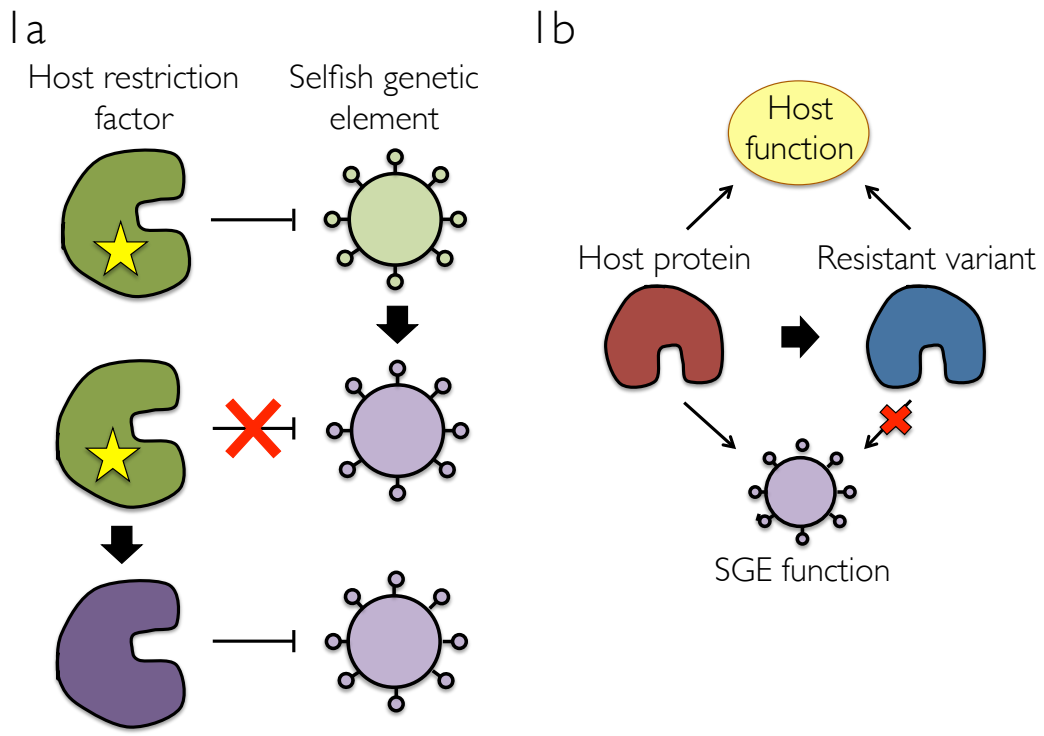

Figure 1.1: Antagonistic coevolution drives molecular arms races 1.1.a) Restriction factors are host immunity proteins that actively target and interfere with SGE function. This antagonistic interaction drives evolution of resistant variants, returning selective pressure to the host to restore immunity. Over time both parties rapidly evolve, driven by the need to "win" this arms race for survival. 1.1.b) Permissivity factors are host housekeeping components that SGEs exploit. Natural selection favors host variants that can escape SGEs while still performing host function; however this may not always be possible.

This ongoing molecular arms race between host and selfish element can drive rapid evolution of genes at the interface of this struggle (Fig 1.1). This rapid evolution (also referred to as positive or diversifying selection) can be quantified as an elevated rate of non-synonymous (amino acid altering, dN) substitutions over synonymous (silent, dS) substitutions that are especially concentrated in domains or specific amino acid residues that lie at the interface between host and parasite proteins[17,18].

Host genes that interact with parasites can be broadly divided into two categories. The first category consists of restriction factors (Fig 1.1.a), which are host-encoded proteins that are dedicated to curb or block parasite success, but do not serve a non-defense role in the host[5]. In such cases, positive selection is driven by an increase in the probability or affinity of host interaction with parasite, which would lower parasite fitness. For example, host variants of restriction factors that better bind and inactivate a parasite are selected for and will increase in frequency within the population. Iterative rounds of host improvement, followed by parasite "escape" leads to signatures of positive selection. A second category consists of permissivity factors (Fig 1.1.b), which encode some host-specific role, but are hijacked by the parasite for its own success. In contrast to restriction factors, parasite interactions with permissivity factors increase parasite fitness but lower host fitness. In many of these cases, we can observe positive selection at the host-parasite interface but here natural selection favors host variants that escape from binding in permissivity factors. However, it is also possible that the housekeeping function of the host factor is so constrained that it cannot permit any changes, making it the perfect target for parasite hijacking[19].

Positive selection is not the only outcome of genetic innovation as a result of arms races between host and SGEs[20]. Frequently, such arms races can give rise to significant gene turnover: restriction factors once duplicated can diversify their anti-parasite repertoire, while loss of permissivity factors could be selected for in other cases. Recombination between paralogs can also create new combinations. Finally, in rare cases, entirely new genes can be coopted into a species' genome (via horizontal gene transfer) as a direct result of conflict with selfish elements. Sometimes this results in a completely new function within the host. One example is syncytin, which was viral in origin, but is now essential for placenta formation in mammals[21]. In other cases, mimicry ensues when hosts steal and deploy parasite genes for defense against

that same SGE, or pathogens mimic host proteins to block host immunity (as is the case with viral K3L - a virally coopted mimic of eIF2alpha)[22].

In some cases it may not be immediately obvious if a genetic element is selfish or is in fact commensal with the rest of the host genome. However, through exploring how a host and genetic element co-evolve, it may become clear if this is in fact an antagonistic relationship[3,16]. Studying how a host genome responds to a genetic element can distinguish between these two possibilities. For example, interactions between symbionts are predicted to result in increased constraint to reflect the interdependency between the two partners, whereas interactions between antagonists are expected to result in increased divergence because interactions that are beneficial to one competitor are detrimental to the other.

Although signatures of innovation are expected in cases of host-parasite interactions, they also provide a means to dissect these interactions. For example, the study of positive selection has provided resolution of host-virus interaction interfaces at single amino acid residue resolution[23]. In other cases, unexpected signatures of positive selection in otherwise essential housekeeping genes suggest that these genes may be involved in genetic conflicts. These genes may be permissivity factors under pressure to escape cooption, or may even be players in novel genetic conflicts. For instance, the unexpected discovery of positive selection in several essential centromeric proteins led to the proposal of the ‘centromere-drive’ hypothesis of selfish centromere transmission during female meiosis in plants and animals[24,25].

In all cases these SGEs shape the evolutionary paths of their hosts. These changes could occur via alterations of allele transmission rates, selective pressure by host fitness costs, or the generation of new alleles. Each of these, in turn, can give rise to newly incompatible genomes[20]. For example, rapid evolution between centromere components has been proposed to give rise

to post-zygotic reproductive isolation between emergent species. Thus, determining the molecular mechanisms by which genetic conflict can give rise to biological diversity and variation within and between species is important for understanding fundamental evolutionary processes.

1.3 BUDDING YEAST: A MODEL SYSTEM TO EXPLORE COEVOLUTION WITH SGES

Despite their ubiquity, it can often be difficult to study genetic conflicts in their native contexts. An ideal system for studying genetic conflict would consist of a host and selfish element that are both genetically tractable, well suited to comparative genomics, and are amply sampled and sequenced. Preferably, this system would be amenable to experimental evolution. Budding yeast and their native selfish elements, particularly the 2-micron plasmid, meet all of these criteria.

Saccharomyces cerevisiae is a powerful species for doing evolution and natural variation studies, as many strains are available to use in the laboratory, and many had been sequenced even at the onset of this project[26–32]. In subsequent years even more *S. cerevisiae* resources have been published (1011 strains and 100 genomes project, for example)[33,34]. In addition, *S. cerevisiae* is arguably one of the best studied model eukaryotes, with many molecular tools and resources available[35]. Combined, this makes *S. cerevisiae* an incredibly well powered system to explore both molecular evolution and functional questions.

Budding yeast species, and in particular *S. cerevisiae,* have 3 main classes of known selfish genetic elements: retrotransposons (Ty elements), the L-A and M “killer” RNA viruses, and 2-micron plasmids[7]. Here I focus on the 2-micron plasmid: in part because this element is a DNA-encoded, non-chromosomally integrated mobile element, which makes it more easily studied by

traditional genetic approaches than either the virus or high-copy, integrated retrotransposons. Furthermore, the 2-micron plasmid has been less well studied in the last 20 years leaving significant opportunities for discovery. For example, while the plasmid's own genes have been well characterized by the Jayaram lab and others, and a short list of host permissivity factors have come to light, we still lack a complete understanding of how host and parasite interact[36–42]. Moreover, until recently, relatively little had been explored in regards to 2-micron plasmid natural diversity within *S. cerevisiae,* as the plasmid had traditionally been omitted from mapping of sequencing reads[43,44]. There have been exciting new developments especially on the sequencing front since this thesis project was initiated.

Excitingly, there exist other increasingly well-sequenced and somewhat experimentally tractable budding yeast species[29,45]. Importantly, some of these other yeast species are also known to harbor their own 2-micron-like plasmids[46–48]. These plasmids have coevolved with their host genome to be stable in the lineages they are natively found in[49]. This observation suggests an ongoing and ancient coevolution, ideal for exploring how antagonistic coevolution has driven changes down multiple lineages and how reproducible or varying these trajectories have been in a broad clade of organisms. Although these diverged plasmids have largely not been explored since their initial characterization in the 1990's, they open an exciting possibility for broad evolutionary studies on both sides of a naturally occurring arms race. However, these known plasmids are found in species sprinkled across the budding yeast clade, with large divergence between host lineages. This divergence makes evolutionary bioinformatics studies challenging or impossible without gathering additional intervening plasmid sequences, which may or may not even exist. If not, how have these other lineages managed to rid themselves of this plasmid? Have they found similar means, or has each lineage solved it uniquely? Taken together, this system is uniquely poised for both broad comparative evolutionary studies and for

in-depth functional molecular work to understand how a host and selfish genetic element shape one another over time.

1.4 THE 2-MICRON PLASMID RELIES ON PLASMID AND HOST GENES FOR ITS PROPAGATION

The best understood aspects of 2-micron biology have been determined primarily through analysis of the plasmid itself[36,39,50,51]. However, *S. cerevisiae* 2-micron plasmids harbor only four protein coding genes of their own (Fig 1.2). This makes 2-micron plasmids obligate genetic parasites, relying on host factors to replicate their DNA and ensure their passage to the next generation of host cells. Although the plasmid was first identified in the 1970s, and had a brief research heyday (where it was exploited for molecular biology tools such as recombinant high copy expression vectors and the repurposing of FLP1 recombinase for genome engineering), we still do not have a complete understanding of how the host and plasmid interact, and indeed how they might be shaping each other's evolutionary trajectory[52–54].

A body of work to date has identified a catalog of host-contributed factors that interact with the plasmid during its normal lifecycle; however this list is likely incomplete, and functional questions remain[36,53,55]. Many of these identified host factors are additionally essential for the host (such as replication machinery, cohesin and members of the sumoylation pathway, among others), suggesting that host permissivity factors exploited by the plasmid may be under tight host functional constraint in spite of selective pressures to escape parasitism[37,40,56,57]. This catch-22 situation may help explain how the plasmid has been such a successful parasite in spite of its 1-3% fitness burden to the host (see section for 1.5 more details) and why it continues to survive in many populations of wild yeasts today[58,59]. Would yeasts that can successfully restrict the plasmid suffer fitness tradeoffs in host functions? Understanding how the plasmid and host interact might predict where this conflict plays out and how a fitness

tradeoffs might arise. Conversely, finding strains or species that have successfully minimized plasmid exploitation might reveal more about typical plasmid lifecycle.

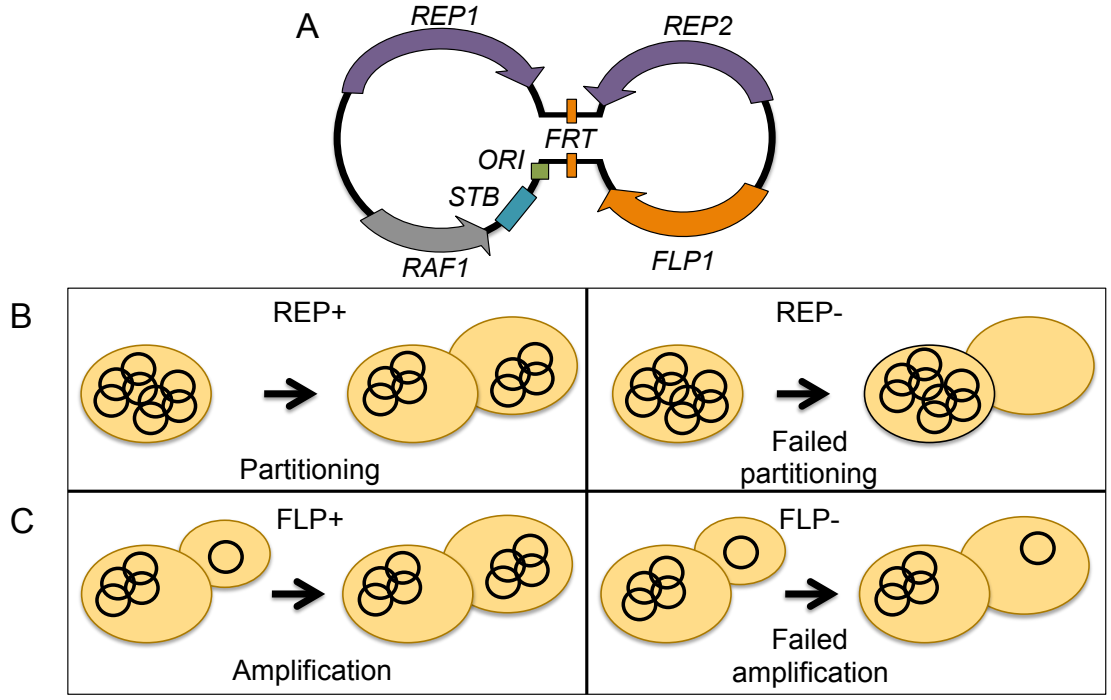

Figure1.2: The 2-micron plasmid and its gene functions 1.2.a) The *S. cerevisiae* 2-micron plasmid has only 4 genes and 4 cis-acting elements. 1.2.b) REP proteins are essential for proper plasmid partitioning into daughter cells. 1.2.c) Following a missegregation event, FLP1 recombinase facilitates copy number amplification of the plasmid during S phase.

1.4.1 *Replication*

The 2-micron plasmid exploits host cellular machinery to replicate its genome. The plasmid origin of replication (or ARS - Autonomously Replicating Sequence) is recognized and licensed (the first step of DNA replication) by the same host licensing machinery as chromosomal origins (Fig 1.2.a), and the plasmid replicates along with the host genome during S phase[56]. Indeed, the 2-micron origin is an early firing origin and since it is a high copy element, it may be unsurprising that extremely high plasmid copy number can lead to replication stress and eventually host cell

cycle arrest[37]. A manifestation of this stress is the observed “nibbled” phenotype, where colonies exhibit frequent sectoring due to plasmid over-abundance. The 2-micron origin ARS consensus sequence (ACS) is identical to some host chromosome origin ACS sequences, although the flanking sequence is unique[60]. Is it possible for some host’s origin licensing machinery to differentiate host from plasmid ARS sequences? If so, host-plasmid conflict may drive divergence in origins of replication over time, even while the origin licensing machinery itself would be under tight functional constraint to maintain host viability.

1.4.2 *Partitioning*

REP1 and *REP2* genes encode DNA binding proteins that, together with the cis-acting plasmid *STB* (‘stability’) locus, ensure high fidelity plasmid segregation to daughter cells (Fig 1.2.b). That REP proteins form a complex with one another and must bind *STB* to achieve proper plasmid partitioning has been well studied, but it is still unknown exactly how this mechanism interacts with the host to achieve equal plasmid partitioning[36,61,62]. Microscopy studies reveal that, although high in copy number, the 2-micron plasmid forms on average ~5 discrete foci at the nuclear periphery[63]. This observation suggests that, although it is present at high copy number within a host’s nucleus, the plasmid effectively segregates as a low copy element. In addition the localization of these foci lead to speculation that these plasmid “bundles” may be tethering to host chromosomes to hitchhike during mitosis (although because budding yeast undergo closed mitosis it is also possible that the plasmid is instead tethering to nuclear membrane associated proteins). Additional work in recent years suggests that the 2-micron plasmid interacts with other components as part of its partitioning process (e.g. Rsc2, Cohesin and microtubules - see Rizvi 2017 review for more complete partitioning overview)[36,38,64–66]. These components, largely identified through biochemical studies, are also essential for host chromosome biology. This makes separating host and plasmid function challenging, and testing of partitioning models non-trivial with traditional genetic and biochemical methods. By exploring

natural variation in plasmid stability and host sequence, I may discover that evolution has already found separation of function mutations. These mutations would alleviate host fitness costs resulting from plasmid exploitation, and facilitate tools to allow for more direct testing of plasmid partitioning models without the use of highly engineered experimental systems.

1.4.3 *Copy number maintenance*

In addition to the partitioning proteins, the 2-micron *FLP1* gene encodes a recombinase that, along with *FRT* sites, ensures plasmid amplification from low copy number following a missegregation event (Fig 1.2.c)[67–70]. Amplification only takes place when the plasmid copy number drops within a cell because the plasmid encoded REP complex also serves to repress *FLP1* when at the stable, high copy number in the host. Additionally, it has been shown that Sumoylation of Flp1 is critical for maintaining plasmid copy number and stable function within the host[40,71,72]. Mutations in the *REP* genes and *FLP1* recombinase have significantly different outcomes for 2-micron stability; the REP proteins are essential for plasmid function, and without them the plasmid is rapidly lost from a host population. However, Flp1 is normally only active when plasmid copy number drops - either due to an earlier replication defect or a missegregation event. As such Flp1 is less critical if other plasmid components are functioning properly, and indeed unless a daughter cell receives no plasmid at all, the 2-micron will continue to persist over generations. This outcome suggests that partitioning functions are more crucial than amplification for 2-micron survival and that interference in plasmid partitioning, rather than copy number suppression, would be the more efficacious host immunity mechanism.

In spite of this body of work, the plasmid has arguably been understudied in the last five decades, in part because of the challenges of disentangling plasmid biology from essential host genes and functions. Many open questions remain, including that of the plasmid's segregation mechanism, as well as the genomic impact of host-plasmid coevolution. By leveraging natural

variation, new avenues of plasmid research are open to pursuit. Naturally occurring alleles that support robust strain growth may be used to explore plasmid function *in vivo*, complementing the synthetic biology and *in vitro* studies being done by others.

1.5 THE 2-MICRON PLASMID: SELFISH OR COMMENSAL?

This minimal plasmid genome composition necessitates that the plasmid parasitize host machinery for survival. Nevertheless, this reliance raises the question of whether 2-micron plasmids are truly SGEs or whether their propagation is better attributed to commensal-like behavior. This question was nicely addressed in the mid-1980s by two seminal studies (Mead *et al*. and Futcher and Cox), which showed that *S. cerevisiae* strains carrying the 2-micron plasmid grew 1-3% more slowly than their isogenic counterparts[58,59]. This fitness cost places a burden on the host, and selective pressure for the host to avoid this parasitism.

As further evidence, many mutant strains are sick in the presence of the 2-micron, but growth can be partially rescued if the plasmid is cured. A classic example of this phenomenon is the *nibbled* phenotype. The mutant *Nib1* showed “nibbled” colonies – due to sectoring as the colonies grew up over time, unless the strain was cured of 2-micron plasmids, and then regular smooth colonies formed. It was later determined that *Nib1* was actually a hypomorphic allele of a ubiquitin E1, *ULP1*, and the nibbled colony phenotype was due to a subset of cells that stalled at the G2-M transition, presumably due to a build up of very high plasmid copy number[37,73]. These and other data suggest the plasmid puts an additional burden on the yeast, both under rapid laboratory growth conditions and in times of stress.

It is also worth noting that there are currently no known conditions in which the 2-micron is beneficial to the host. Unlike bacterial plasmids, the 2-micron is not observed to integrate other genes that may be beneficial to the host (such as resistance genes common in prokaryotic

plasmids)[74]. And, unlike the "killer" viruses also found in budding yeasts, which encode a toxin and antitoxin system where yeasts that lose the virus die, there does not appear to be an addiction mechanism associated with the 2-micron plasmid[7,75]. Rather than context-specific benefits, it may be that the plasmid continues to survive by being difficult for the host to escape. First, with the *FLP1* mediated self-repairing copy number amplification system, only a complete lack of plasmid transmission to daughter cells will successfully eliminate the plasmid. Second, the host encoded proteins that are known to interact with the 2-micron largely are essential for host function as well, so mutations affecting plasmid use of these proteins may impair normal host functions as well. Third, the plasmid can be easily reintroduced into populations through mating. Fourth, the plasmid is subject to non-Mendelian inheritance, with all progeny inheriting the 2-micron through meiosis, even when only one haploid parent of a cross was initially infected. Finally, even if rare meiotic plasmid loss were to take place during sporulation, given the relative high frequency of within-ascus mating in budding yeasts, the plasmid would likely be rapidly reintroduced into a cell upon sister-spore diploidization. Taken together, I posit that unless a strain were to evolve means to *heritably* restrict the plasmid, it would be unsurprising that the lifestyle and constant threat of reintroduction make the 2-micron a successful and effective parasite. Furthermore, it would be interesting to understand if *S. cerevisiae* domestication has influenced 2-micron spread within the species, and if this human interference is shaping coevolutionary patterns differently relative to undomesticated yeast species.

Here I argue that, in spite of its prevalence within *S. cerevisiae* as a species, the plasmid is in fact a selfish parasite that is well optimized to exploit its natural host, rather than being beneficial in some contexts. It has been seen that the plasmid is maintained in natural populations and has coevolved to be high copy number and stable in the lineages they are naturally found in[49]. When a plasmid is introduced into a non-native context (such as introduction into a new species) the plasmid loses stability and copy number[49]. This lineage

specificity suggests plasmid and host cell coevolution, either through positive selection or host drift followed by plasmid mutations to restore stability. There may indeed have been rounds of host escape, followed by plasmid counter adaptation to their native host, which has driven this divergence of plasmid sequence. This provides an opportunity to study the genetic antagonism between budding yeast host genomes and the 2-micron SGE.

1.6 LEVERAGING NATURAL VARIATION TO STUDY THE CONFLICT BETWEEN *S. CEREVISIAE* AND 2-MICRON PLASMIDS

Although we have known about the plasmid for more than 40 years, and it exists in arguably one of the best characterized eukaryotic model organisms, we still do not fully understand the mechanisms by which the plasmid and host interact at the mechanistic level, or the evolutionary repercussions of these interactions. This model host-parasite system provides an exciting venue for broad comparative evolutionary studies as well as deep mechanistic exploration using the bevy of molecular and genetics tools available in budding yeast. This combination is especially exciting in the context of a naturally occurring and ecologically relevant host-parasite arms race, where both players are experimentally tractable. It is rare that host-parasite interactions can be mechanistically studied in the lab in their endogenous contexts, with many systems requiring non-native hosts (e.g. mouse models of human pathogens, or tissue culture rather than whole-organism work), recombinant parasites (e.g. viral minireplicon systems rather than intact infectious particles), or purely observational work (such as epidemiological studies) in natural systems.

Based on the premise that 2-micron plasmids are indeed SGEs, and that they appear to have adapted to their respective host genomes, I chose to study if there are cases within *S. cerevisiae* in which host cells have evolved to repress 2-micron transmission. I hypothesized that some natural yeast isolates may have found molecular means to limit plasmid exploitation.

Although we do not know the full contingent of host encoded permissivity factors or restriction factors for the 2-micron plasmid in budding yeasts, taking an unbiased screening and genetic mapping approach could help determine genetic variants that limit plasmid success in some strain backgrounds. This approach would be informative in understanding more about basic plasmid biology, but also in understanding the evolutionary routes a host might take to successfully combat an otherwise successful parasite. Taking this natural variation strategy allows us to explore the molecular changes that were effective at plasmid restriction while allowing the host to remain viable and competitive enough to survive outside of laboratory conditions. Leveraging natural variants allows for the possibility of studying essential genes (through naturally arising polymorphisms) in a way that a genetic knock-out screen can not. This approach is helpful in exploring what tradeoffs might exist for the solutions reached for limiting plasmid success, and is of particular relevance since many of the known 2-micron permissivity factors are essential to the host as well.

To study the 2-micron plasmid and natural host variants, I first needed a robust assay to quantify 2-micron stability in different genetic backgrounds in a robust assay that could be scaled for a genetic screen. In Chapter 2 I present such an assay which significantly improves beyond traditional methods to study high copy number plasmids such as 2-microns. Next, I wished to assess whether I could detect natural variation in 2-micron propagation among wild strains of *S. cerevisiae*. I present the results of this approach in Chapter 3. In Chapter 4, I present our efforts to use one wild strain that represses 2-micron propagation to map the genetic determinants of 2-micron stability. Finally, in Chapter 5, I discuss my overall findings and the future prospects to dissect this fascinating example of genetic conflict further, using new resources and alternative methods for this study.

Chapter 2. SCAMPR: A SINGLE-CELL ASSAY FOR MEASURING PLASMID RETENTION

2.1 INTRODUCTION

Plasmids are mobile genetic elements: extrachromosomal, circular pieces of DNA that are vertically transmitted in microbial populations as well as through sex. Some plasmids in bacteria, for example, act as infectious agents by increasing copy number within a host and spreading through the host population through bacterial conjugation[74]. These genetic elements often confer fitness costs, but also context-specific benefits to the host; for instance a plasmid that imposes a cost to bacterial fitness may nonetheless contain an antibiotic resistance gene that is useful to the host cell upon exposure to specific antimicrobials[74]. Still other plasmids propagate within the microbial population in spite of fitness costs to the host, even without any known benefit to the host.

Plasmid stability (or plasmid loss) assays allow for the temporal study of plasmid maintenance in populations of cells, leading to a greater understanding of how these mobile elements propagate within populations and impact the hosts they occupy. Observing if and how plasmids are lost over time can provide a readout of plasmid stability but also reveal aspects of host cell function, especially in cases where plasmids rely on host machinery for their replication or propagation. For example, artificial reporter plasmids can be used to assess whether a particular host cell pathway is functioning normally. Indeed, host genomic mutations leading to plasmid loss were previously used to identify and characterize essential replication genes in budding yeast[76]. Cells with hypomorphic mutations in their replication machinery would lose plasmids more rapidly than cells with normal replication machinery. In this case, plasmid instability was a convenient phenotype to monitor function of essential host genes within a cell.

Understanding plasmid host range and stability is important from both biological and biotechnological perspectives. For example, plasmids may be used for developing new vector tools for use in emerging model systems, such as facilitating protein overexpression studies[6877]. In these cases, plasmid stability assays can be used to explore the function of vectors in different host backgrounds to understand whether the plasmids being developed have a broad or narrow range of host use. Similarly, understanding the natural dynamics of plasmid loss can provide a better understanding of spread and stability of antimicrobial resistance within and between bacterial species in the environment. Without a true measure of plasmid stability within host species, it would be impossible to evaluate the role of plasmid mobility between species. Understanding the biological basis of host-plasmid compatibility could thus also help prevent spread of, and increase host resistance to, parasitic mobile genetic elements and the cargo that they carry.

Plasmids can be lost from dividing cells through different means; understanding these loss dynamics may reveal the underlying molecular mechanisms. For example, an exogenous plasmid in a new host may not be replicated well. This underreplication would lead to a decrease in copy number each generation, with plasmids gradually being lost from the population of host cells (Fig 2.1.a). Alternatively, plasmids could be missegregated when host cells divide, with some daughter cells not receiving any plasmid, and other daughter cells harboring those additional copies (Fig 2.1.b). Under this scenario, plasmids would replicate appropriately but be unable to exit the ‘mother cell’ resulting in a precipitous loss in the population as a whole. Being able to experimentally determine the difference between these loss dynamics can reveal what host functions plasmids rely on for their stability. This insight could be leveraged to improve expression vector stability in new models, or exploited to limit the spread of parasitic elements.

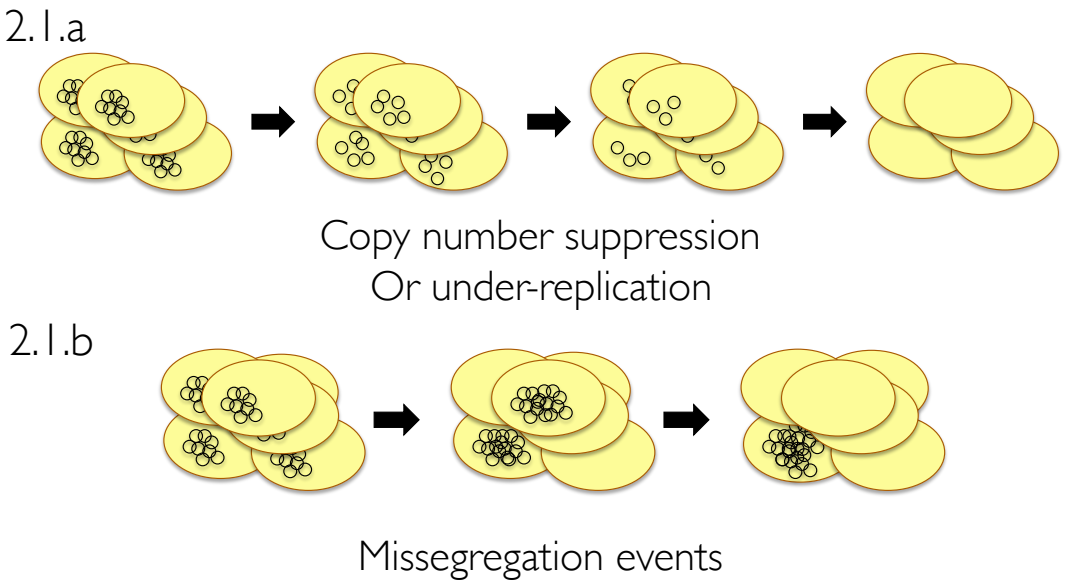

Figure 2.1: Plasmid loss dynamics. 2.1.a) Consistent copy number decrease, by under replication or degradation mechanisms, leads to steady loss of plasmid. 2.1.b) Uneven plasmid partitioning during cell division leads to population heterogeneity and plasmid loss.

2.2 TRADITIONAL ASSAYS TO MEASURE PLASMID STABILITY

Previously, plasmid loss dynamics have primarily been measured by Minichromosome Maintenance (MCM) assays, qPCR and Southern blot based assays. Briefly, the MCM assay assesses plasmid occupancy within a population over time by plating samples of a population of cells harboring plasmids with selectable markers on both selective and non-selective media at regular intervals during non-selective growth of the population (Fig 2.2.a). The proportion of cells that have completely lost the plasmid (and therefore can no longer grow on selective plates) is counted to estimate a loss rate for the plasmid[76]. This method is extremely useful to measure the plasmid loss dynamics of low or single copy plasmids in host cells. However, it is unable to detect variations in copy number within cells, as usually only a single copy of a selectable marker is required to get viable cell growth. In the case of high copy number

plasmids, substantial variation in plasmid stability and loss could potentially go undetected by this assay.

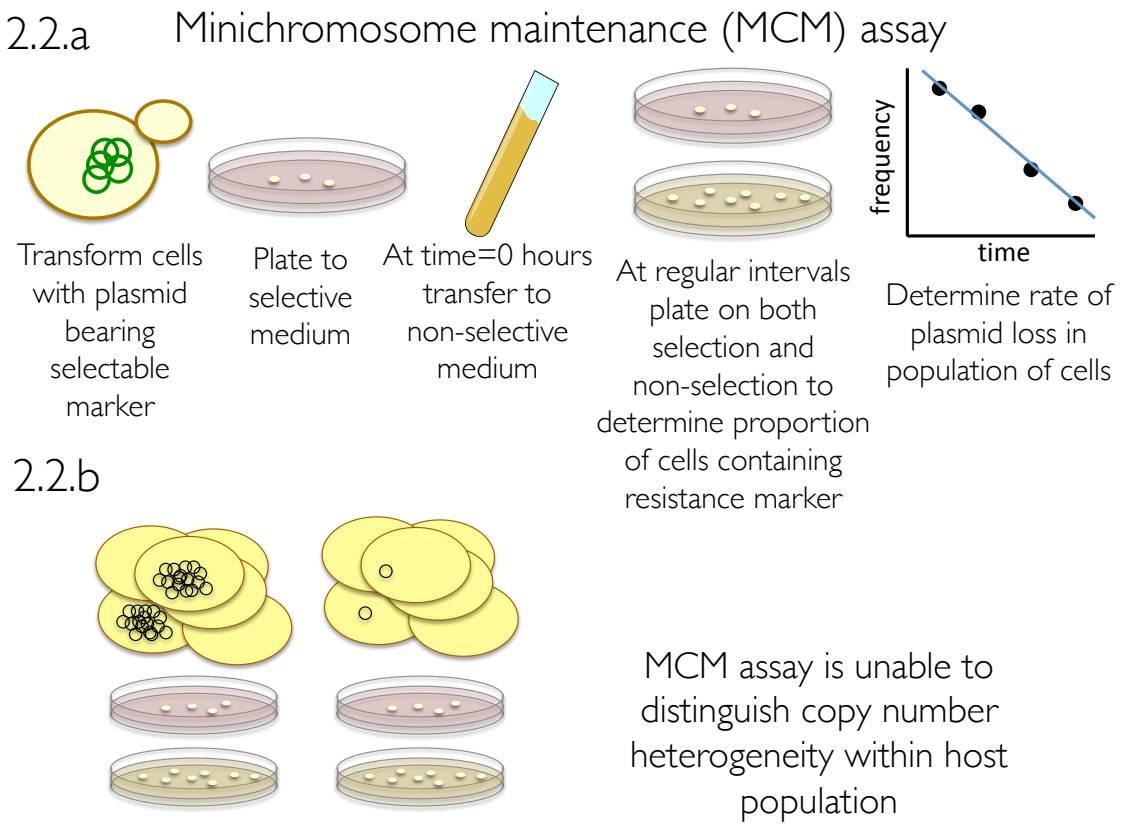

Figure 2.2: Minichromosome Maintenance Assay Fig 2.2.a) Assay workflow Fig 2.2.b) MCM assay cannot detect plasmid copy number heterogeneity, only total-loss events

Other assays such as qPCR and Southern blots can be more useful than the MCM assay to measure high copy number plasmid changes (Fig 2.3.a)[78]. In both cases DNA is prepared from a large number of cells, and the copy number of the plasmid is assessed relative to the genomic DNA. Conveniently, they have a significant advantage over MCM assays in that the plasmid being assayed does not require genetic manipulation to introduce selective markers. However, even these assays can only assess the copy number of the plasmid averaged across an entire population of host cells. By only measuring the mean copy number, these assays cannot reveal

any information about population heterogeneity or copy number distribution (Fig 2.3.b). Again, strains with very different plasmid inheritance dynamics (such as unequal partitioning events) would be undetectable by these assays.

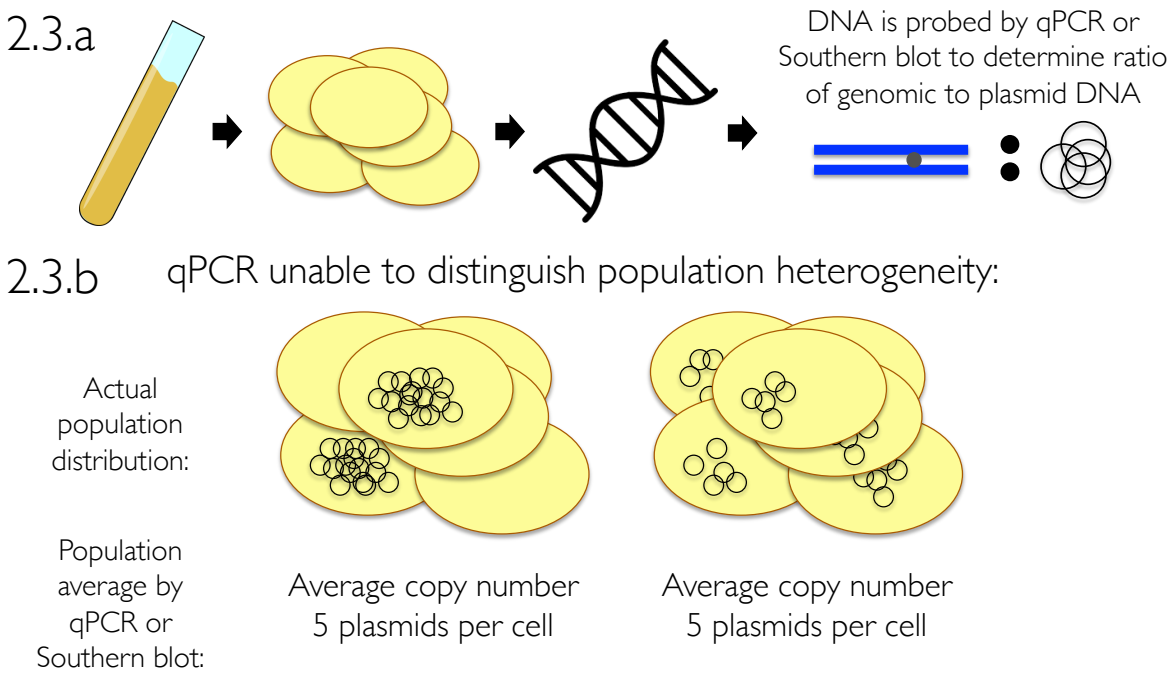

Figure 2.3: qPCR and Southern blots compare genomic and plasmid DNA abundance within a population of cells. 2.3.a) DNA is prepared and relative stoichiometry of plasmid and genomic DNA is assessed. 2.3.b) qPCR and Southern blotting are unable to detect population heterogeneity; changes in copy number are averaged over the population of cells.

In some instances, the MCM and qPCR based assays could together provide information about both copy number and the frequency of total plasmid loss events within a host population. However, even in combination, these methods lack the resolution to really understand the distribution or variability of plasmid abundance within a host population. Thus, there are limitations to both types of traditional plasmid loss protocols to measure the stability of high-copy plasmids. A better understanding of the dynamics of plasmid loss would require single-cell measurements that assess plasmid copy number and distribution across a population. Ideally,

such single-cell measurements should be able to scale to thousands or millions of cells to enable statistically supported comparisons.

2.3 2-MICRON PLASMIDS AS A MODEL TO STUDY HIGH-COPY NUMBER PLASMIDS

To compare between different assays of high copy plasmid stability, I decided to focus on 2-micron plasmids, which are naturally occurring, high copy number, extrachromosomal elements found in budding yeast. The life cycle of these plasmids in particular make them an exciting case study for exploring how mechanisms of loss can give rise to different plasmid loss dynamics. These plasmids have no known benefit to their host, and in fact confer a fitness cost to the cells that carry them. Nevertheless they are successful within *S. cerevisiae*, with most strains harboring this parasite, including laboratory strains. Although the plasmid was first identified in the 1970s and some aspects of plasmid biology have been well characterized, there remain open questions about how this plasmid and host interact.

The 2-micron plasmid employs two different mechanisms to ensure its stability in host cells. Although they are present at high copy number (~60 in haploid cells in lab strains of *S. cerevisiae*), 2-micron plasmids segregate in a small number (1-5) of discrete foci when the host undergoes mitosis. These plasmid “bundles” are stoichiometrically replicated along with the host genome in S phase, then sister bundles are segregated to the daughter cells during mitosis. Unlike yeast chromosomes that encode centromeres to ensure stable segregation, the plasmid does not enucleate a kinetochore. Instead, the plasmid encodes two protein components REP1 and REP2, which act in concert with a cis-acting “stability” locus to mediate segregation. Mutations in these plasmid partitioning factors can have catastrophic consequences on plasmid stability. These components cannot act alone; they require host factors to function normally, but

we currently lack a comprehensive understanding of this partitioning mechanism or which host components are minimally required.

Fascinatingly, following a missegregation event, 2-micron plasmids have the means to recover high copy number even when only a fraction of the plasmids are inherited. This mechanism relies on the plasmid encoded Flp1 recombinase and FRT sites that ensure that the plasmid undergoes over-replication by orchestrating recombination during S phase of the cell cycle. Although this mechanism plays an important role in plasmid stability, mutations in the Flp1 recombinase have less dramatic consequences on plasmid stability as these mutations only bear consequence if a missegregation or under replication event take place to begin with.

Despite their complicated lifestyle and intricate involvement with host machinery, a long history of research in 2-micron plasmids makes these an ideal system to test existing and new assays for plasmid stability within a population. In particular, previous discovery of host and plasmid mutations known to affect plasmid stability and copy number are ideal to predictably influence plasmid stability and assess if different assays can reliably measure these changes.

2.4 Results and Discussion: Devising a single-cell plasmid assay to measure 2-micron stability

Motivated by a better understanding of 2-micron plasmid stability and dynamics in budding yeast populations, I developed a high throughput assay to measure plasmid stability that scales well for large numbers of strains and also captures population heterogeneity of plasmid copy number. This assay utilizes flow cytometry paired with a fluorescent reporter plasmid to capture single-cell plasmid information, which facilitates higher throughput experiments and can reveal information about underlying plasmid loss dynamics (Fig 2.4.a). By utilizing flow cytometry to

assess single cells I can simultaneously measure two properties captured by more traditional assays. For example, analogous to the MCM assay, my assay allows me to detect and count the number of total plasmid loss events by measuring proportion of GFP negative cells (Fig 2.4.b). At the same time, analogous to the qPCR/Southern assay, I can measure the mean plasmid copy number based on mean GFP intensity (Fig 2.4.c). Notably, however, by performing this assay at the single-cell scale, I not only infer the population average but also identify the population distribution surrounding that average (Fig 2.5), revealing the inherent heterogeneity underlying plasmid loss. Based on the fact that my assay is designed to measure plasmid stability at the single-cell level, I will refer to it as SCAMPR (for single-cell assay for measurement of plasmid retention) from here on.

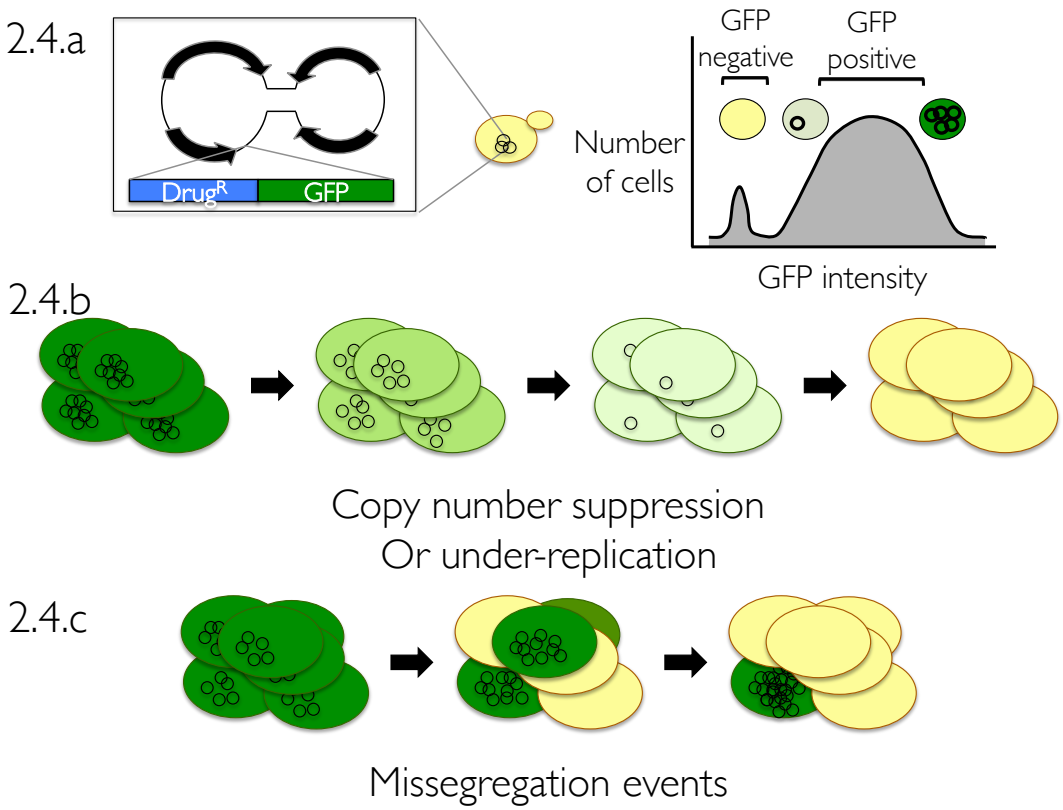

Figure 2.4: Fluorescence as a marker for plasmid presence and copy number. 2.4.a) a reporter plasmid expressing GFP and a drug marker can be used to select and screen for plasmid in yeast cells. Single cell measurements can be assessed by flow cytometry to explore

population distribution and heterogeneity. 2.4.b) Plasmid copy number changes are detectable by changes in fluorescence intensity. 2.4.c) Missegregation events result in GFP negative cells. Both parameters can be measured for a population at the same time.

I first wished to assess whether GFP expression can be used to track plasmid stability *in vivo*. I chose not to utilize recombinant 2-micron episomal vectors (yEP) that have been traditionally used for yeast over-expression studies for several reasons. First, these vectors contain only a small fraction of the endogenous 2-micron sequence (the origin of replication and STB locus) in an otherwise bacterial cloning vector backbone. This construction means that the proteins required for yEP vector functionality must be provided in trans by the endogenous 2-micron plasmid itself. For SCAMPR, I wanted to make a self-contained reporter for broader utility in different strain backgrounds and that would more closely parallel endogenous plasmid function. I note here that these vectors are somewhat less stable than the endogenous plasmid; I attribute this reduced stability to increased fitness cost from constitutively expressing a reporter cassette; however, I can not rule out the possibility that transcription of this cassette may impact plasmid stability through other means.

Instead, to test the propagation of 2-micron plasmids, I constructed a GFP-2-micron reporter that contains a cassette with both a selectable marker (G418 resistance) and a screenable eGFP marker, each under a constitutive promoter[79]. To avoid disruption of the plasmid's endogenous replication and segregation machinery, the cassette was integrated into the 2-micron sequence found in BY4741 (type A) at a restriction site previously published to be tolerant of insertions up to 3.8 kb DNA[77]. I also avoided the use of bacterial cloning sequences, as they would have increased the size of the insertion and potentially decreased plasmid copy number and stability. The GFP-2-micron plasmid was created by Gibson assembly directly into

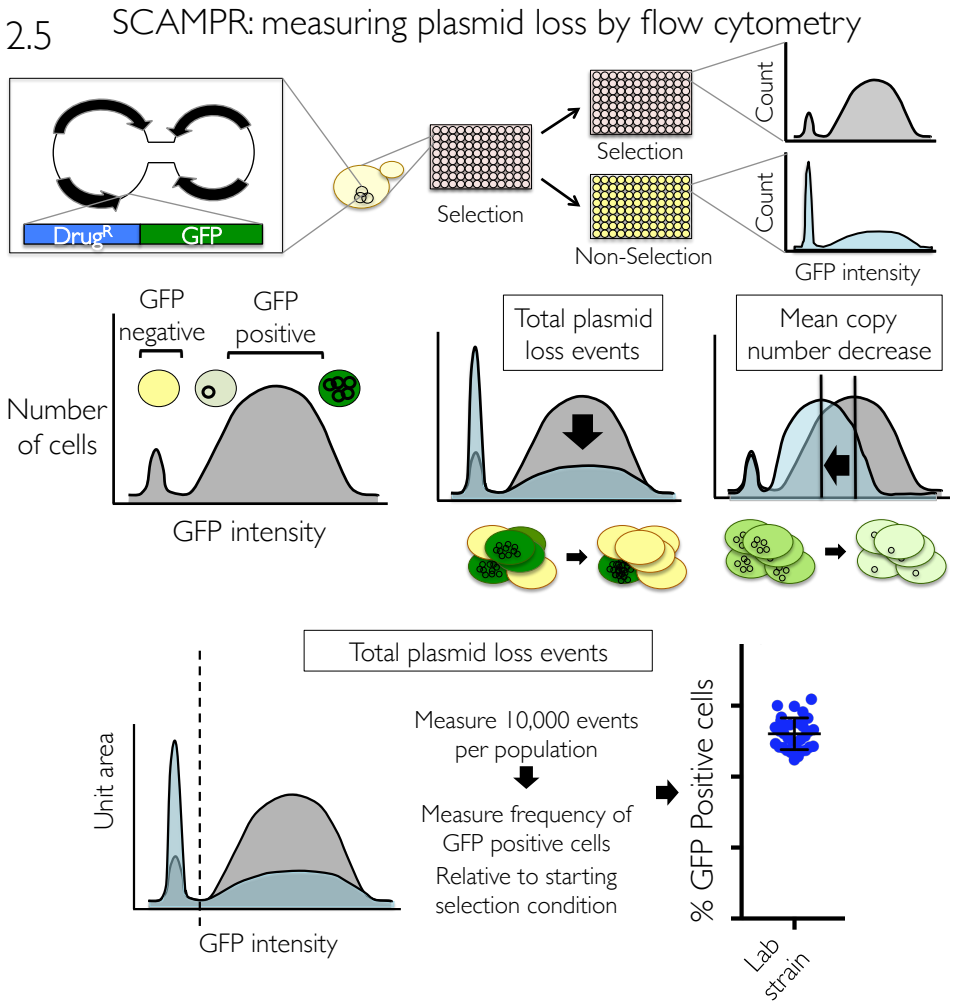

Figure 2.5: SCAMPR - A **S**ingle-**C**ell **A**ssay for **M**easuring **P**lasmid **R**etention by flow cytometry

otherwise plasmid-less yeast strains (also referred to a *cir*^{0}). These lab strains, BY4741 (MATa haploid) and BY4742 (MATα haploid), were first cured of their endogenous plasmids by methods previously published by the Gartenberg lab[80]. This step avoided multiple plasmid genotypes within a strain background that could have led to plasmid competition or recombination.

2.5 COMPARING THE STABILITY OF 2-MICRON PLASMIDS VIA DIFFERENT ASSAYS

With the help of Paula Levan, I tested the stability of the 2-micron plasmids in four budding yeast strains: a wild type lab strain which supports plasmid function well, two lab strain mutant backgrounds described by others as losing 2-micron plasmids more rapidly, and a natural isolate that does not natively harbor the 2-micron plasmid. The mutants lose 2-micron plasmids either by failing to support plasmid copy number (dominant acting *flp1* step-arrest mutant) or segregation (*rsc2* null mutant) respectively. The natural variant *S. cerevisiae* strain that does not have the 2-micron plasmid is Y9, also called Ragi (see Chapters 3 and 4 for further discussion and studies of this natural isolate). These strains were chosen to validate the SCAMPR assay because we anticipated they would have differing plasmid stabilities, and that the underlying mechanisms and population dynamics of plasmid loss would be different between these strain backgrounds.

A previously characterized mutant of flp1 recombinase is known to create double stranded DNA breaks at the FRT sites, but fails to facilitate recombination as the normally functioning protein would[80]. This dominant-acting mutation was introduced under an inducible promoter into BY4742 cells containing the GFP-2-micron reporter plasmid using the same plasmid as described in the original paper[80]. This *flp1* mutant competes with the reporter plasmid's functional copy of FLP1 and decreases the copy number of 2-micron within cells. *RSC2* is a member of the RSC chromatin remodeling complex and while *rsc2* null mutants are viable, they show increased frequency of missegregation events for both 2-micron and chromosomes[64,81]. Others report only a slight fitness defect, in *rsc2* null mutants, so I was surprised to observe our strains have a significant growth defect[64]. The prior plasmid stability studies of *rsc2* were done in W303 strain backgrounds, which in my hands have a slower doubling time than BY4742, but I

still was not anticipating this difference based on the mutant. This was true for all recovered isolates of this genotype, regardless of presence or absence of the 2-micron plasmid. A more recent publication suggests that *rsc2* mutants may be somewhat respiratory deficient which would explain my decreased growth observations[82].

I tested the stability of the GFP-2-micron plasmid in multiple strains by SCAMPR, my new single-cell high throughput plasmid loss assay, as well as by Minichromosome Maintenance (MCM) assay (Fig 2.2.a) to provide a more appropriate direct comparison[76]. I first measured plasmid stability using the SCAMPR assay (Fig 2.5). The experimental set up for SCAMPR is similar to that of the MCM assay to begin with, but measures plasmid loss by fluorescence of individual cells rather than through colony counts on selective media. Briefly, yeast cells bearing plasmid are grown under selective conditions until time=0. At that point, cultures are split into both selective and non-selective media and grown at 30C with shaking. At 24 hours (~12 generations), I measure GFP intensity for >10,000 single cells per sample (selective and non-selective conditions) by flow cytometry. The distributions for the two conditions are then directly compared. I performed this assay for BY4742 wild type, *rsc2* null and *flp1* mutant cells, and Ragi wild type cells.

By SCAMPR I find that the laboratory strain maintains the fluorescent reporter plasmid more stably than Ragi (Fig 2.6). I explore this observation and the underlying biological basis for this difference further in Chapters 3 and 4. I also note here that the *rcs2* null showed less 2-micron plasmid loss than we had expected based on prior publications (Fig 2.7.a). This may be occurring for a number of reasons, however I suspect that the cells are not undergoing the same number of generations, even by extending the incubation time to 48 hours and/or including additional passaging steps. I observe that these strains seem to stop growing at a substantially lower OD, in addition to having significantly longer doubling times even during log

phase growth. I base this on observations of fewer cells after overnight growth and less turbid growth even after 48 hours of incubation relative to the other strains. The prior publication reporting 2-micron loss did not account for generation time differences, reporting only a “minor” fitness defect of the *rsc2* null mutation in their hands. I was surprised to note this difference in growth, and am currently working to correct for this generation time difference and lower saturation density.

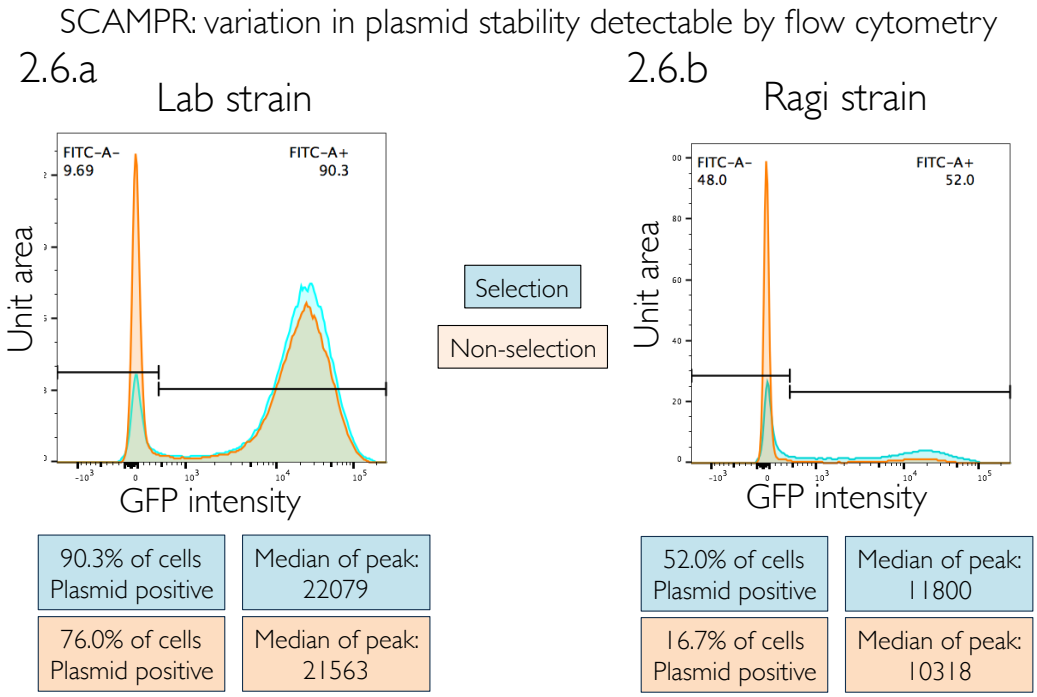

Figure 2.6: SCAMPR comparable to MCM in data quality for total-loss events, while collecting additional data more rapidly

It is notable that even under selection, and without galactose induction it appears that the step arrest mutant already shows a significant 2-micron copy number decrease (Fig 2.7.b). When under SC-MSG -ura, glucose, +G418 conditions I expected the cells would not express the galactose-inducible mutant FLP protein, and would resemble the WT in similar conditions. However, I see the median GFP positive peak is lower than the WT strain with the same GFP-2-micron reporter: WT = 8524 IU, *rsc2* = 4913, *flp1* mutant = 6488, and indeed the distribution of

GFP is substantially different in this strain as well, suggesting a difference in copy number even under uninduced conditions. The original paper creating the step arrest plasmid did not quantify plasmid loss in the induced and uninduced conditions - only reporting a qualitative increase in 2-micron loss in strains patched on solid galactose media conditions. At this time I am unsure whether my data reflect the same results the original authors observed, or if my strain has a different phenotype. I suspect it is somehow constitutively expressing the step-arrest flp protein in spite of the galactose-inducible promoter construct. I am using their construct without any additional modifications, however some background or unsequenced vector mutations could still underlie this unexpected change. I would find it somewhat surprising if a novel mutation created a constitutively active promoter in just my strain, rather than a dead promoter. Instead, I suspect this change in plasmid distribution reflects both the underlying biology of the plasmid loss (mutant lines lose plasmid), but also inherent assay bias. Reassuringly, SCAMPR as an assay reveals this copy number decrease, regardless of underlying biological cause.

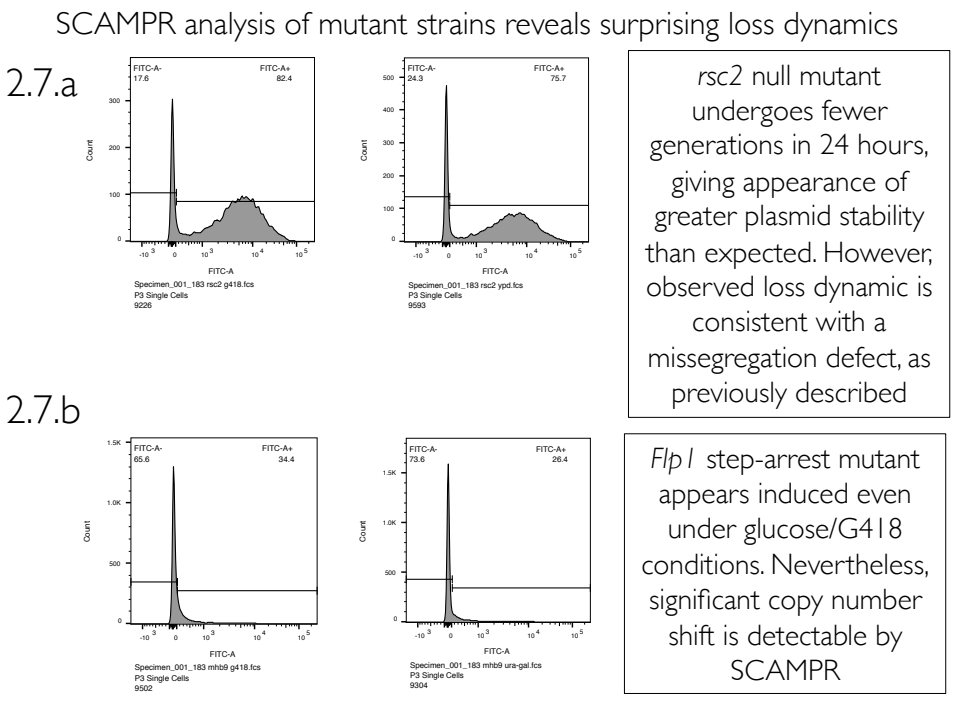

Figure 2.7: SCAMPR analysis suggests surprising population dynamics of plasmid loss in mutant strains

BY4742 and Ragi wild type cells were also tested by MCM assay to validate the SCAMPR results. I find that the flow cytometry-based SCAMPR assay captures similar total-loss event data as the MCM assay for Ragi and BY4742, but with the added benefit of simultaneously reporting changes in the distribution of plasmids within the GFP positive population of cells as well all in a single experiment (Fig 2.8). In the MCM assay, briefly, strains containing the reporter plasmid are initially grown under G418 selection to ensure 2-micron plasmid maintenance. Then at time=0, cultures are transferred into liquid media with shaking, but without drug selection for 12 generations (or ~24 hours). After 24 hours, these cultures were diluted in PBS and plated on rich medium either with or without G418 selection at multiple dilutions, targeting 30-300 CFU per plate. Plates were incubated for 2 days, then colonies were counted to determine what fraction of the population were G418 positive and therefore contained the 2-micron plasmid at the final time point relative to the initial time point. I find that the MCM assay data corroborate results I observed by SCAMPR for both BY4742 as well as the Ragi strain, showing that SCAMPR is not only a suitable replacement for the MCM assay, but also is faster and captures additional data. The *flp* and *rsc2* mutant lines are currently being validated by MCM as well, but were not prioritized for assay validation due to the concerns outlined above.

Further work to optimize rsc2 null growth conditions could help reconcile our data with what has been published previously. We anticipate that ongoing MCM results will match the SCAMPR data well, if not the previously described plasmid instability result. Regardless, this comparison between MCM and SCAMPR assay results for the BY wild type as well as Ragi show that SCAMPR is an equally effective method for monitoring plasmid stability in a population of cells.

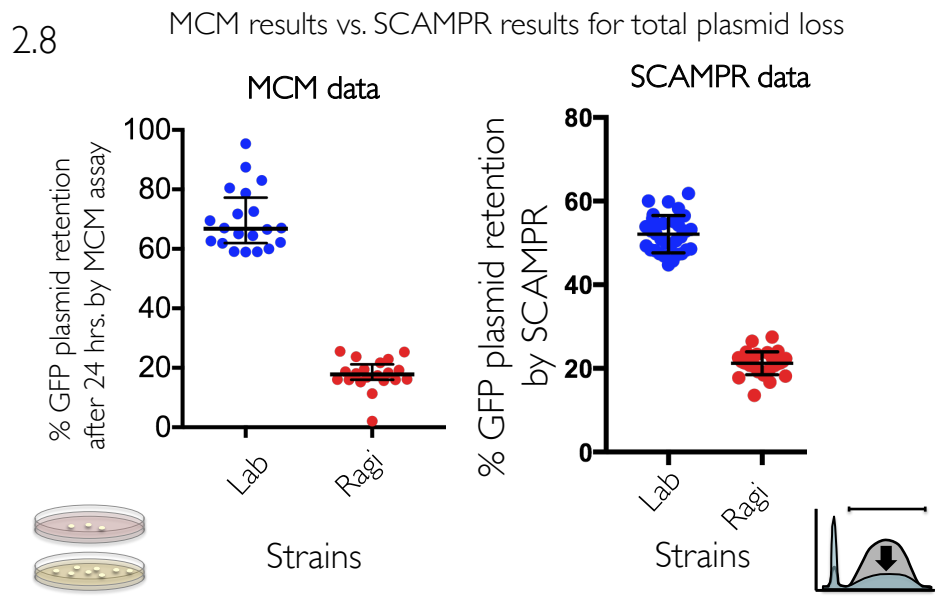

Figure 2.8: SCAMPR comparable to MCM in data quality for total-loss events, while collecting additional data more rapidly

To ensure that the GFP-G418 reporter cassette was well expressed and stable in my strain backgrounds (irrespective of plasmid dynamics), I checked eGFP expression of a single copy, chromosomally integrated cassette in lab strain, and Ragi strain backgrounds (Fig 2.9). The median GFP intensity was the same between strain backgrounds (WT BY4742= 1307 IU, Ragi = 1332), indicating equal expression; I interpret this as evidence of equal promoter efficiency and/or protein turnover in both strains. Furthermore, nearly all cells in both strain backgrounds were GFP positive (>99.9%), indicating that the eGFP cassette is not especially prone to silencing in either of these strain backgrounds. These data suggest that GFP expression from the cassette is both stable and consistent across strain backgrounds, and gave me confidence that loss, or decrease, of GFP in these strains could be used as a proxy for plasmid loss.

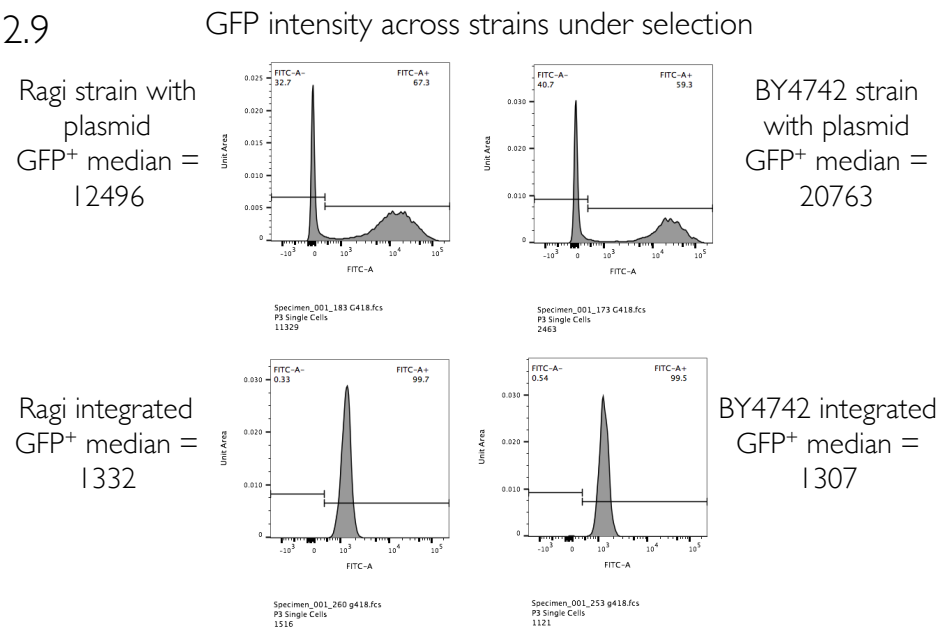

Figure 2.9: Fluorescent reporter cassette shows stable and equal expression in different strain backgrounds, and even at high copy number the distribution of GFP intensity appears not to saturate.

Also critical to the efficacy of the assay is the premise that GFP expression levels correlate well with DNA copy number in budding yeast. I wondered if GFP intensity could also be used as a proxy for plasmid copy number. Indeed, other studies have used GFP expression to track copy number and genome structural changes in yeast[83–86]. If plasmid copy number can be estimated from GFP intensity, then flow cytometry would facilitate both plasmid loss and copy number assessment, and the heterogeneity thereof, all in a single assay. A related concern is whether GFP signals could saturate at high plasmid copy number. At least in this 2-micron study system saturation did not seem to be the case. Indeed, the observed GFP intensity curves fall in normal distributions when at high copy number (Fig 2.9) which suggests that the GFP signal has not saturated. Additionally, even if the GFP signal was saturated in some cells, it would occur at high copy number levels and most studies of plasmid loss and stability are more likely to be focused on the lower copy number range.

2.6 FUTURE DIRECTIONS: REVISITING FITNESS ESTIMATES OF 2-MICRON PLASMIDS USING AN IMPROVED ASSAY

The greater resolution of the plasmid stability of 2-micron plasmids using the SCAMPR assay could allow me to revisit some of the classic experiments in the field that first elucidated that 2-micron plasmids impose a 1-3% fitness cost to budding yeast, identifying them as pathogens rather than symbionts or commensals. In these experiments, 2-micron plasmids containing yeast strains (*cir*$^{+}$) were competed against plasmid negative (*cir*^{0}) cells[58,59]. Over time, the *cir*^{0} cells were found to increase in frequency relative to the *cir*$^{+}$ cells. To confirm if this increase was due to fitness cost, the study authors seeded competitions at different strain ratios, and compared their data to predictive models of host fitness cost. These experiments allowed the study authors to conclude that loss of plasmids conferred a 1-3% fitness improvement to *cir*^{0} strains thereby allowing them to outcompete *cir*$^{+}$ cells still harboring the 2-micron plasmids.

However, these previous studies lacked the ability to couple lineage tracking with plasmid stability. Therefore, the study authors could not conclude whether the increase in cells was solely due to a fitness increase of *cir*^{0} strains, or whether a small rate loss of 2-micron plasmids from *cir*$^{+}$ strains led to this over-representation (Fig 2.10.a). By pairing lineage-tracking with the SCAMPR assay, I could unambiguously measure how this increase in *cir*^{0} cells arises. By empirically determining how new plasmid-free cells arise I can not only distinguish between two models (fitness vs. stability) but also determine to what degree the two mechanisms contribute to the increase in *cir*^{0} cells.

Therefore Paula Levan and I created genetically marked *cir*^{0} and *cir*$^{+}$ strains to distinguish them from each other to facilitate lineage tracking[8788]. Briefly, dsRed and CFP and Cerulean

expressing cassettes were integrated into the host chromosome of both *cir^0 and cir$^+$* cells. To understand whether the frequency increase of the *cir^0* cells occurred solely on account of their higher fitness I assessed which lineage the increase in *cir^0* were arising from. We chose two blue proteins in hopes that one might give better separation of signal from the eGFP channel. I found that the eGFP marker selected for our plasmid is so bright that it bleeds into the blue channels, regardless of CFP or Cerulean (AmCyan or Pacific blue lasers). This coupled with how dim these proteins are in general made them poor choices for lineage tracking in this case. I had anticipated YFP would also be too similar to our reporter 2-micron, so we did not use this marker. Instead a far-red protein might be suitable as a replacement in future studies. DsRed marked cells were distinct from the eGFP, however, so for our initial experiments we elected to track one lineage with dsRed, and the other lineage without an additional marker, then measured what proportion of the GFP negative population is red or not marked. This is less ideal because unmarked cells could arise due to protein loss or marker silencing as well, however I see that infrequently in 24 hours of growth of the red marked cells alone. Additional fitness cost of the dsRed marker can also be separated from plasmid stability by doing this experiment in both directions (fig. 2.10.b).

I infer the 2-micron status of the cells by separately following the GFP-marked 2-micron plasmids as before by SCAMPR. Competitive growth assays are set up between cells with and without the plasmid as was done in previous studies by Mead *et al.* and Futcher and Cox, at 50:50 ratio initially with cells that are actively dividing to avoid difference in lag phase from skewing results. These initial experiments are currently ongoing; the pilot data reveled that the dsRed negative and positive cell distributions are overlapping when comparing between different strains. This can perhaps be solved by change in the flow cytometer settings to get better separation of data at the lower fluorescence intensity in the red channel while still

preserving the large range of sensitivity for the eGFP channel. These experiments are currently underway.

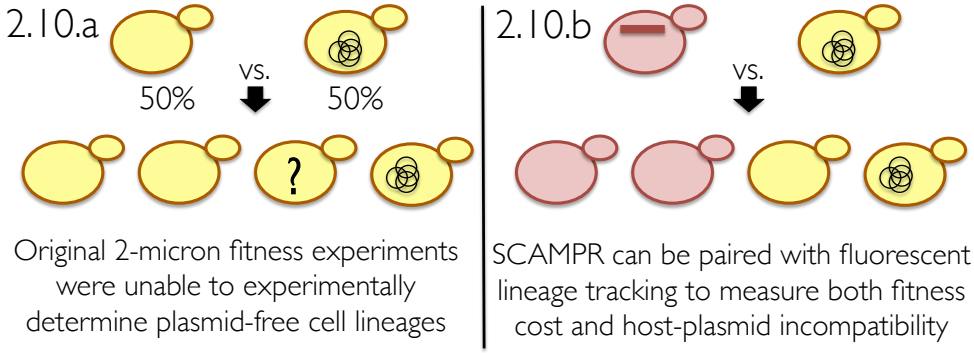

Figure 2.10: SCAMPR allows for measurement of both plasmid fitness cost as well as host-plasmid incompatibility that lead to plasmid loss. 2.10.a) Original experiments could not differentiate between fitness cost and plasmid-loss experimentally, so data were fit to models of each possibility. 2.10.b) SCAMPR allows for lineage tracking and plasmid measurements in the same cells.

2.7 Discussion

Here, I present a new assay, SCAMPR, which leverages the power of fluorometric measurements to measure plasmid stability at the single-cell level. My assay is able to detect more nuanced changes in plasmid copy number and stability, while providing the increased power as traditional assays to measure the frequency of complete loss events. SCAMPR evaluates both plasmid copy number and complete-loss events in a single assay; a comparative approach would require implementation of both MCM and qPCR assays and even then, the latter would not be able to reveal the true extent of heterogeneity in plasmid stability.

While developing the SCAMPR assay, I considered whether the perdurance of GFP proteins might mask DNA loss in this assay, if the protein were present in the cells long after the plasmid had been eliminated. This sort of phenotypic lag time might obscure DNA loss, or make it challenging to detect small differences in DNA stability between different strains. While this may

be true of drug resistance markers as well, those are the current state of the art in existing assays. In my experimental system I did not observe an increased lag; the proportion of cells without plasmid at 24 hours (15 generations) were similar by MCM assay (based on G418 selection) and by GFP flow cytometry measurement assay. If GFP phenotypic lag time differed greatly from G418 lag time I would expect different values between the two measurements, but I do not observe this difference, demonstrating GFP in this system is an equally responsive marker for DNA stability as the current standard. However, this may not be the case in all systems. If DNA and protein stability differences lead to poor correlation between plasmid loss and GFP loss, this method could be extended to use kinetically destabilized fluorescent proteins. These destabilized proteins would minimize the phenotypic lag caused by proteins that are more stabile than the DNA that encodes them. This modification could help extend assay sensitivity in comparing strains with more similar loss rates or particularly unstable plasmids.

Compared to traditional methods of assessing plasmid stability, SCAMPR has the added benefit of much higher throughput, larger sample sizes and faster turn around time than traditional plasmid loss assays. This difference is a non-trivial. For instance, flow cytometry facilitated measuring 10,000+ events per strain background, whereas CFU are limited to ~300 per plate. In addition, multiple samples – both replicates and multiple strain backgrounds - can be easily run in 96- or 384-well plates for the flow cytometry assay. This greatly improves throughput relative to plating assays, as each well can be directly measured. In the MCM assay, one must plate multiple dilutions of the same sample to ensure CFU counts are between 30-300 per plate. This range requires some expectation of how frequent plasmid loss is to begin with. Whereas eight 96-well plates (4 selective and 4 nonselective condition plates each) are easily run in a single experimental round in the SCAMPR assay, given access to a flow cytometer, the same number of samples would require nearly 1600 agar plates by MCM assay, assuming only two plating dilutions per sample. Furthermore, since there is no colony outgrowth required in the SCAMPR

assay, phenotyping in “real time” by flow cytometry makes the assay faster overall and also avoids confounding plasmid loss during the early microcolony forming generations. Finally, the SCAMPR assay requires less hands-on time than the MCM assay.

The low throughput and time intensive nature of the MCM assay means that the MCM assay cannot feasibly be used to phenotype large genetic crosses or strain panels, since only a few strains can be easily assayed in parallel. This restriction limits its utility for certain types of studies of the genetic basis of plasmid-related phenotypes. Additionally, traditional MCM or qPCR plasmid stability assays do not account for the distribution of plasmid copy number within a population – data that can be informative about underlying plasmid loss mechanisms. Inferring *how* a plasmid is lost in a population, such as change in copy number distribution, could generate hypotheses about the mechanisms underlying this loss (underreplication vs. plasmid non-segregation for example). Both of these impediments can be overcome using the SCAMPR assay.

Moreover, the SCAMPR assay may be expanded for doing comparative fitness experiments. This extension will be useful for disentangling the fitness costs plasmids impose on their hosts from their inherent rate of instability. I am currently revisiting iconic experiments on the fitness costs of 2-micron plasmids, hoping to independently measure the effects of both fitness cost and plasmid instability on the increased recovery of plasmid-free cells during competitive growth experiments. Simultaneous tracking of host cells and plasmids will provide an exciting new avenue of research to explore host-plasmid interactions in multiple experimental systems.

2.8 METHODS

2.8.1 *Strain growth and construction*

Unless otherwise noted, yeast strain construction and growth was performed by standard yeast methods including lithium acetate transformation, and growth at 30C[89]. Strains used in this work are listed in Table 1.1. The reporter GFP-2-micron plasmid was Gibson assembled directly into *S. cerevisiae* using NEBuilder HiFi DNA Assembly Master Mix (product E2621), then harvested by Zymoresearch Zymoprep Yeast miniprep kit (D2004) and retransformed to ensure clonality.

2.8.2 *MCM assay*

MCM assays were performed as others have previously published, however samples were taken at only two time points, and as such were reported as changes in frequency rather than an estimated rate[76]. (I believe more than two measurements are needed to accurately estimate rate, however this significantly increased workload, and this better matches the SCAMPR assay workflow as well.) Samples were plated at multiple dilutions to ensure between 30-300 CFU per plate were available. All strains containing the reporter plasmid were grown under G418 selection to ensure 2-micron plasmid maintenance prior to the start of the assay. At time = 0 hours, cultures were transferred into liquid media with shaking, but without drug selection for 24 hours. After 24 hours, cultures were diluted in PBS and plated on YPD either with or without G418 selection at multiple dilutions, targeting 30-300 CFU per plate. Plates were incubated for 2 days, then colonies were manually counted to determine what fraction of the population were G418 positive. Calculations were based on whichever dilution gave a countable (30-300 CFU) plate. Multiple replicates (at least 8) were done for each strain to establish variability.

2.8.3 *SCAMPR*

SCAMPR assays were performed as described in the body of Chapter 2. The sample setup was similar to MCM assay, however cells were measured by flow cytometry directly at time = 24 hours. A BD Canto-2 cytometer was used to collect cell data, with FlowJo software used for subsequent data analysis. For ease of scheduling the selection and non-selection conditions were both measured at time = 24 hours. This was validated by measuring 24 hour growth in selection as compared to time = 0 growth in selection. The same population distribution was observed for both, so I determined these could be measured interchangeably. I attribute this population distribution as inherent to a strain; the number of cells without plasmid while under selection is constant as it reflects the phenotypic lag for that genetic background. Cells that lose the plasmid die under selection, but are constantly generated as well. Repeated passaging could give rise to mutant escaper populations, however this is unlikely to occur in only 24 hours of growth, and no loss of drug sensitivity (or gain in plasmid stability) was ever observed for a strain background in this study.

Chapter 3. CHAPTER 3: NATURAL VARIATION IN 2-MICRON STABILITY WITHIN *S. CEREVISIAE*

3.1 INTRODUCTION

The 2-micron plasmid is a selfish mobile genetic element found only in budding yeast species. These DNA plasmids are high copy, extrachromosomal elements. Like other selfish elements the 2-micron both relies on the host for its own propagation, but confers a fitness cost to their host in the process. In spite of this burden, the 2-micron plasmid is found in most wild and domesticated *Saccharomyces cerevisiae* strains. Additionally, 2-micron-like plasmids are also found in multiple diverged species scattered throughout the phylogeny of budding yeasts, where these plasmids have evolved to be stable in only their native host species. These findings suggest that despite pressure on the host to lose them, 2-micron plasmids have adapted to maintain stability in their host lineages.

Although the 2-micron plasmid was first described in the early 1970s, and the plasmid encoded genes have been well described, it is still not fully understood how host and plasmid interact, or the evolutionary ramifications of these interactions. Of the host factors that have been found to interact with the plasmid, many perform essential host functions as well. This essentiality has made decoupling host and plasmid functions experimentally difficult by traditional genetics and biochemical methods. For example, a genetic screen for loss of plasmid stability would only be able to reveal non-essential host genes, although many processes that the plasmid relies on the host to perform (e.g., DNA replication, chromosome segregation) are likely to be essential for the host cell as well.

To survive in the host, plasmids must replicate their DNA and ensure their passage to the next generation. However, the 2-micron plasmid's minimal genome requires that they co-opt host cellular machinery (permissivity factors) to accomplish processes essential for their survival. If the host were able to avoid this permissivity factor cooption, while still fulfilling host cellular functions, the host could successfully evade parasitism. However, these genes' evolutionary capacity may be under tight functional constraint given their essential function to the host. If permissivity factor escape is not possible, the host might instead evolve a restriction factor: a host encoded element whose primary function is to actively interrupt normal plasmid function. Cellular immunity factors have been largely uncharacterized in *S. cerevisiae*, in spite of their long history as a popular model eukaryote. The few factors that have been characterized have focused on the Ty retrotransposons that populate yeast genomes, although more recent analyses have begun to explore how the host components identify and restrict the dsRNA genomes of "killer" viruses[90–95].

Here I sought to explore inherent differences within the *S. cerevisiae* species for 2-micron propagation. I reasoned that since *S. cerevisiae* and 2-micron plasmids are locked in an antagonistic relationship, perhaps I could leverage genetic variation within species to identify whether there was phenotypic variation with respect to 2-micron maintenance. This approach would be complementary to the traditional biochemical and genetic approaches taken to date, and could facilitate study of plasmid-interacting cellular factors that have been otherwise challenging to work with in the lab. For example, even though host-essential permissivity factors would be undetectable in a genetic screen of null mutants, it is possible that natural variants of these essential genes have arisen that are able to separate host and plasmid function. Ongoing antagonistic coevolution between host and plasmid means that natural selection would have favored the fixation of such variants at least in some isolates of the species.

Until recently, sequencing efforts in budding yeast have mostly ignored the 2-micron plasmid, although this is changing with more recent sequencing efforts. These studies have suggested that the majority of *S. cerevisiae* strains harbor 2-micron plasmids but this may not be the case in other closely related species. The finding that some *S. cerevisiae* strains lack 2-micron plasmids could reflect true genetic variation for 2-micron propagation or this could represent stochastic loss. Fortunately, SCAMPR (Chapter two) allows us to distinguish between these possibilities. Here, I describe my discovery of natural *Saccharomyces cerevisiae* isolates that have evolved genetic means to rapidly lose 2-micron plasmids. The basis of this resistance appears to be dominant and heritable. My findings provide strong support for the 2-micron as a parasite model. Furthermore, it provides a framework to study the genetic architecture underlying 2-micron stability and propagation in budding yeast.

3.2 RESULTS: WILD YEAST STRAINS VARY IN THEIR CARRIAGE OF 2-MICRON PLASMIDS

For my approach to work I needed to carry out two steps. First, I needed to identify strains that naturally do not harbor the 2-micron plasmid. Second, I had to test whether this natural absence of 2-micron plasmids represents a heritable trait rather than a stochastic loss event. In addition, I assessed whether the basis for 2-micron resistance was dominant or recessive, suggesting the role of restriction versus permissivity factors, respectively.

Because the 2-micron confers a fitness cost on the cells that harbor them, I wondered if any natural *S. cerevisiae* isolates had evolved means to restrict the plasmid. To explore this question I searched for natural yeast isolates that do not contain the 2-micron plasmid. To identify plasmid negative strains I screened a panel of 58 natural *S. cerevisiae* isolates, kindly provided by Justin Fay, for the presence of the endogenous 2-micron. I identified three isolates that in my hands do not contain the 2-micron plasmid (Table 1 - strain table). Initially I screened

Ragi, Oak and Palm wine strains for plasmid presence by PCR (Fig 3.1.a). Briefly, primers were designed to probe each half of the plasmid, generating fragments from the coding regions of an essential plasmid gene (*Rep1*) and the most well-conserved plasmid gene (*Flp1*) respectively. However, PCR is sensitive to polymorphism, so isolates that tested plasmid-negative by both PCR reactions were further tested by Southern blot, to confirm they were plasmid-negative (Fig 3.1.b) rather than harboring polymorphisms that interrupted the primer binding. Finally, the strains found to be plasmid negative were also whole genome sequenced. This approach eliminated concerns of extremely divergent plasmids in these strains. Whole genome sequencing and *de novo* genome assemblies confirmed my prior data that these 3 isolates indeed do not contain endogenous DNA plasmids.

Table 1: Strain table of isolates checked for endogenous plasmid.

Strain Key	Alias	Alternate	Mating_type	From	Comments	Location	2-micron status (by PCR)	confirmed by Southern
YMH1	BY4741	YAD1	A	Aimee Dudley	lab strain haploid, known 2-micron positive		positive	
YMH2	BY4742	YAD147	alpha	Aimee Dudley	lab strain haploid, known 2-micron positive		positive	
YMH3	FY4	YAD145	A	Aimee Dudley	lab strain haploid, S288c lineage, suspected 2-micron positive		positive	
YMH4	FY5	YAD146	alpha	Aimee Dudley	lab strain haploid, S288c lineage, suspected 2-micron positive		positive	
YMH5	G170	F1441, YAD448	A	Tim Galitski	mating type tester halo assay		positive	
YMH6	G4	GY184, YAD449	alpha	Tim Galitski	mating type tester halo assay		positive	
YMH7	Sigma1278b,	YO352	A	Tim Galitski	from Fink lab		positive	
YMH8	Sigma1278b,	YO353	alpha	Tim Galitski	from Fink lab		positive	
YMH9	M1	A1	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH10	M2	A2	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH11	M3	B4	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH12	M4	B5	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH13	M5	B8	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH14	M6	B9	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH15	M7	C6	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH16	M8	C7	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH17	M9	C8	a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH18	B6	Zymoflore VL3	a/alpha diploid	Barbara Dunn via Justin Fay	Zymoflore VL3	France	positive	
YMH19	NRRL y2411	Y8	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	Wine	Turkey	positive	
YMH20	NRRL y390	Y1	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	Mushrooms		positive	
YMH21	NRRL y5997	Y9	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	ragi	Java, Indonesia	positive	
YMH22	NRRL yb427	Y2	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	Rum	Trinidad	positive	
YMH23	NRRL y1438	Y3	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	Palm Wine	Africa	positive	
YMH24	NRRL y7567	Y10	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	Coconut	Philippines	positive	
YMH25	NRRL y1532	Y4	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	Fruit	Indonesia	positive	
YMH26	NRRL y1546	Y5	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	Wine	West Africa	positive	
YMH27	NRRL y11846	Y11	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	alpechin	Spain	positive	
YMH28	NRRL yb1952	Y6	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay		French Guiana	positive	
YMH29	M22		a/alpha diploid	R. Mortimer via Justin Fay	Vineyard	Italy	positive	
YMH30	NRRL y2190	Y7	a/alpha diploid	C. Kurtzman, ARS Culture Collection via Justin Fay	D. psuedoobscura	Yosemite, CA	positive	
YMH31	PR	Pasteur Red	a/alpha diploid	Red Star via Justin Fay		Paris	positive	
YMH32	CDB	Cote Des Blanc	a/alpha diploid	Red Star via Justin Fay		Germany	positive	
YMH33	MR	Montrachet	a/alpha diploid	Red Star via Justin Fay		US	positive	
YMH34	YPS163		a/alpha diploid	P. Sniegowski via Justin Fay	Oak Exudate	PA, USA	positive	
YMH35	PW5		a/alpha diploid	Justin Fay	Palm wine, Nigeria	Raphia palm wine, Aba, Abia state, Nigeria, 2002.	positive	
YMH36	YJM421		a/alpha diploid	Justin Fay	Clinical, US	Isolated from from ascites fluid, pre-1994.	positive	
YMH37	YPS1009		a/alpha diploid	Justin Fay	Oak tree, US	Oak exudate, Mettler Woods, NJ, 2000.	negative	yes
YMH38	IL-01		a/alpha diploid	Justin Fay	Soil, US	Soil sample, Cahokia, IL, 2003.	positive	
YMH39	T7		a/alpha diploid	Justin Fay	Oak tree, US	Tree exudate, Babler State Park, MO, 2003.	positive	
YMH40	NC-02		a/alpha diploid	Justin Fay	Oak tree, US	Tree exudate, Smoky Mountains, NC, 2003.	positive	
YMH41	YJM440		a/alpha diploid	J. McCusker via Justin Fay	Clinical isolate		positive	
YMH42	I14		a/alpha diploid	Justin Fay	Vineyard, Italy	Vineyard soil sample, Petina, Italy, 2002.	positive	
YMH43	Y12	NRRL y12633	a/alpha diploid	Justin Fay	Palm wine, Africa	Palm wine, Ivory cost, Africa., pre-1981.	negative	yes
YMH44	Y10	NRRL y7567	a	Justin Fay	Coconut, Philippines	Pre-1973	positive	
YMH45	Y9	NRRL y5997	a/alpha diploid	Justin Fay	Ragi, Indonesia	Ragi, Java, Indonesia, pre-1962.	negative	yes
YMH46	UC5	UCD612	a	Justin Fay	Sake, Japan	Sene Sake, Kurashi, Japan, pre-1974.	positive	
YMH47	YJM428		a/alpha diploid	Justin Fay	Clinical, US	Isolated from from paracentesis fluid, pre-1994.	positive	
YMH48	YJM436		a/alpha diploid	J. McCusker via Justin Fay	Clinical isolate		positive	
YMH49	YJM421		a/alpha diploid	Justin Fay	Clinical, US	Isolated from from ascites fluid, pre-1994.	positive	
YMH50	CLIB215		a/alpha diploid	Justin Fay	Baker, New Zealand	1994	positive	
YMH51	T73		a/alpha diploid	Justin Fay	Wine, Spain	Red wine of Monastrel grape in fermentation stage, Alicante, Spain, 1987.	positive	
YMH52	CBS7960		alpha	Justin Fay	Sugar cane, Brazil	Factory producing ethanol from cane-sugar syrup, Sao Paulo, Brazil.	positive	
YMH53	YJM269		a/alpha diploid	Justin Fay	Wine, Europe	Blauer Portugieser grapes, 1954.	positive	
YMH54	YJM320		a/alpha diploid	Justin Fay	Clinical, US	Isolated from from blood, pre-1994.	positive	
YMH55	YJM280		a/alpha diploid	Justin Fay	Clinical, US	Isolated from from peritoneal fluid, pre-1994.	positive	
YMH56	YJM326		a/alpha diploid	Justin Fay	Clinical, US	Pre-1994	positive	
YMH57	CLIB324		a/alpha diploid	Justin Fay	Baker, Vietnam	Saigon, Vietnam, 1996.	positive	
YMH58	YJM653		a/alpha diploid	Justin Fay	Clinical	Isolated from brochoalveolar lavage.	positive	
YMH65	DBY7730		a	Ivan Liachkb, Dunham lab			positive	
YMH66	DBY7442		alpha	Ivan Liachkb, Dunham lab			positive	

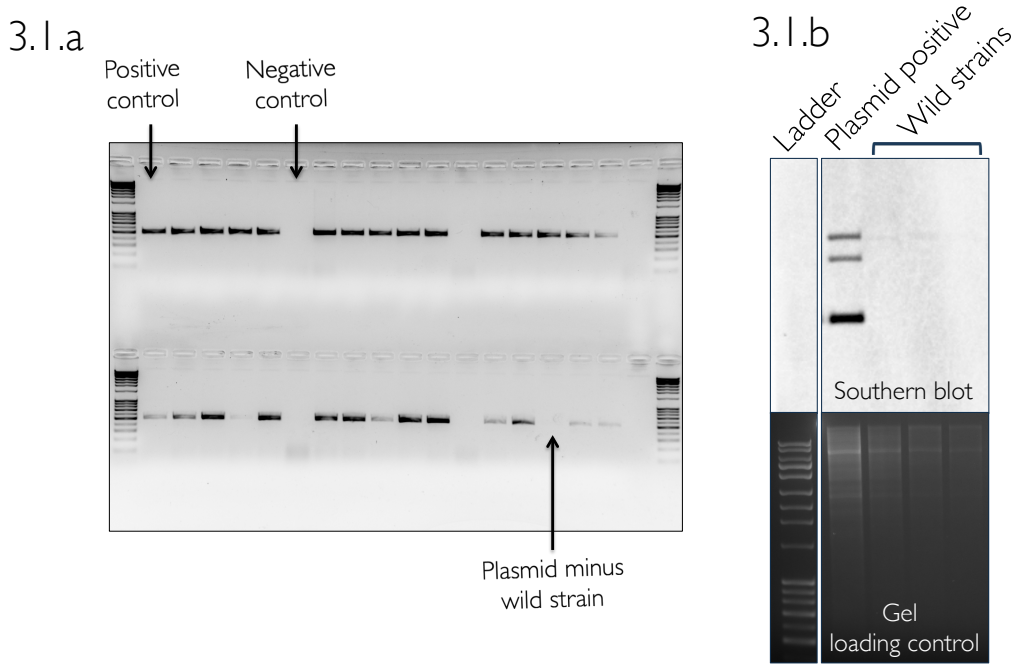

Figure 3.1: Screening for 2-micron-negative strains reveals three candidates. 3.1.a) Strains were screened initially by PCR 3.1.b) Plasmid-negative candidates were verified by Southern blot, and eventually by WGS

3.3 NATURAL VARIATION IN 2-MICRON SUSCEPTIBILITY IS A HERITABLE TRAIT

Since others have found that the 2-micron confers a fitness cost to the host cell, with laboratory strains growing 1-3% better without the plasmid, I wondered if 2-micron absence in these natural isolates is a heritable trait. I hypothesized that these strains may have evolved means to heritably restrict the parasitic plasmid, which would confer a benefit to the host cell. Alternatively, it was possible that these strains had stochastically lost the plasmid previously, and due to their subsequent handling they had not been reinfected. Natural selection would favor isolates that have lost the 2-micron, because of the fitness cost imposed on the host by the plasmid.

Heritable means for host restriction of the 2-micron would be the most efficient means to prevent spread and reintroduction of the plasmid into a population. Because the plasmid is

easily reintroduced through sex, exhibits non-Mendelian inheritance and actively promotes its own segregation and high copy number, stochastic loss may not be sufficient to purge it from a host population. To test if 2-micron plasmid loss in these 3 natural isolates is a heritable trait, I utilized traditional plasmid loss assays as well as SCAMPR (Chapter 2). Briefly, I reintroduced a reporter plasmid into these natural isolates, then measured if the plasmid was lost once again. To accomplish this, I transformed the candidate strains with a GFP expressing and G418 resistant recombinant 2-micron plasmid. The reporter 2-micron was also introduced into the BY4742 lab strain as a control for a plasmid-permissive strain background. I first looked at comparative plasmid stability between strains by colony sectoring assay and was excited to see the striking difference in sectoring between the lab strain and the wild isolates (Fig 3.2.a). These data suggest that these natural isolates have evolved a heritable means for plasmid loss.

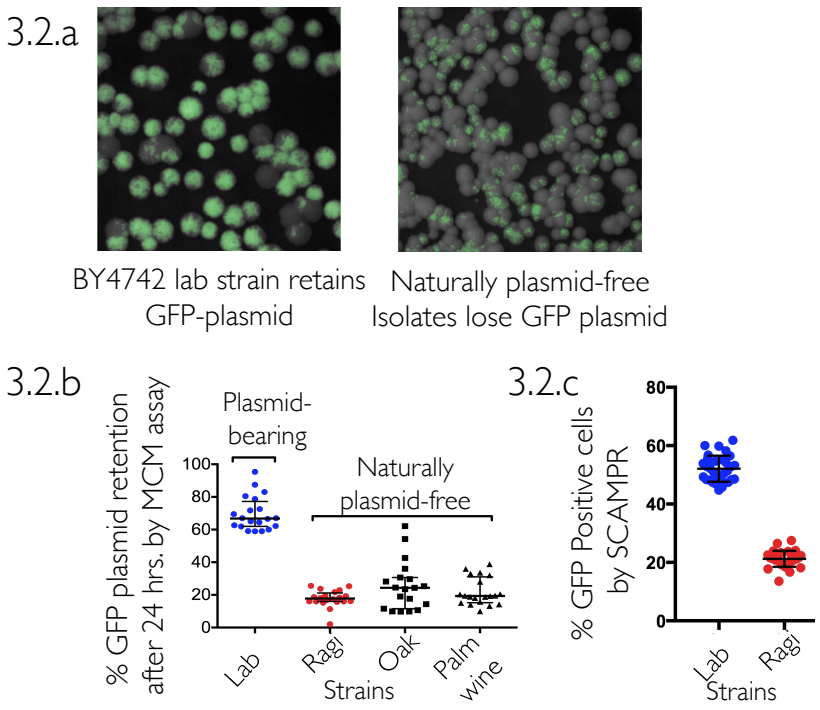

Figure 3.2: 2-micron is less stable in naturally plasmid-negative *S. cerevisiae* isolates. Relative plasmid stability was assessed by 3.2.a) colony sectoring, 3.2.b) MCM assay and 3.2.c) SCAMPR.

Because the colony sectoring assay is qualitative, I next wanted to quantify the difference in plasmid stability in these different backgrounds. I initially measured plasmid loss by MCM assay (Fig 3.2.b), as well as confirming those data by SCAMPR (Fig 3.2.c and Chapter 2). These data confirm that the natural isolates lose the reporter plasmid substantially more quickly than the permissive laboratory strain. Data were normalized to the starting frequency of plasmid-negative cells (even under selection a subpopulation of GFP negative cells persist due to phenotypic lag). At T=24 hours, the naturally plasmid-free isolates maintained substantially less plasmid than the lab strain (Fig 3.2.b), further supporting my observation that these natural isolates heritably lose the 2-micron plasmid. The haploid Ragi strain maintained plasmid in only ~5% of the population on average between replicate experiments (normalized here to 20% to account for the starting frequency of plasmid negative cells under selection), while the lab strain maintained ~60% of the population (~70% normalized). The other 2 strains showed similar loss patterns, with the NJ oak strain exhibiting more variability between replicates than the other two strains. Interestingly, Y9 and Y12 (Ragi and palm wine) have been shown to be closely related. These strains cluster with sake strains and indeed share haplotypes with one another.

This striking plasmid loss data, by colony sectoring, MCM and SCAMPR confirm that the 2-micron plasmid loss trait is both heritable and highly reproducible in these natural isolates.

3.4 RAPID PLASMID LOSS IS A PHENOTYPICALLY DOMINANT TRAIT

I next wondered if this trait is genetically recessive or dominant. I hypothesized that if this plasmid loss is due to mutations within a host permissivity factor, that plasmid loss would likely be recessive or show an intermediate phenotype in a heterozygous diploid (because the other allele would confer plasmid stability). Alternatively, if plasmid loss is due to a host-encoded plasmid restriction factor, this would likely be a dominant trait. To distinguish between these possibilities I compared plasmid stability in homozygous diploids as well as heterozygous diploids created by crossing the BY lab strain to natural isolate haploids. Excitingly, the

heterozygous diploid strains rapidly lost the plasmid, similar to the homozygous natural isolates, suggesting a dominant plasmid loss trait (Fig 3.3). While these surprising data support a restriction factor hypothesis, they do not test this directly nor confirm the presence of a restriction factor. However, they are a tantalizing hint of underlying genetic architecture and are informative for genetic mapping strategies.

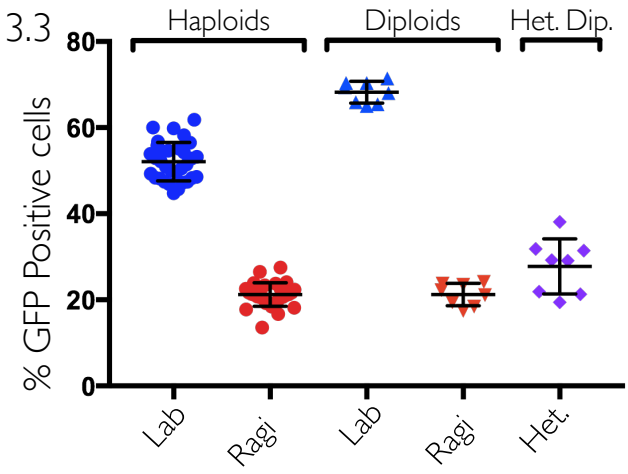

Figure 3.3: Rapid 2-micron loss consistent with a dominant trait in Ragi

3.5 Future directions: Rapid plasmid loss driven by genetic incompatibility, increased fitness cost or both?

Finally, I wanted to explore whether this increase in plasmid loss rate is because of higher fitness costs in some strain backgrounds, or if it is due to host-plasmid incompatibility. Although these natural strain backgrounds have very similar doubling times in rich media as the lab strain, both with and without the plasmid, fitness differences less than ~5% can be challenging to detect without doing competitive growth assays. Previously, others have shown that in lab strains the 2-micron confers a 1-3% fitness cost to the host, but it is unclear if this is variable in different strain backgrounds. As described in Chapter 2, these results did not distinguish the difference between cells that arose from parent lineages with or without the plasmid. To address this ambiguity I wanted to combine the flow cytometry loss assay SCAMPR with lineage tracking (Fig 3.4.a) to measure how this increase in plasmid negative cells arise. These data could be

compared to the same type of data from the laboratory strain to assess natural variation in these contributing drivers of plasmid loss (Chapter 2.6, future directions). These data could further test between two possible reasons for plasmid instability in Ragi. If Ragi is less fit when the plasmid is present this could suggest that plasmid loss is selected for because of a synthetic growth defect between the host genome and 2-micron plasmid. Alternatively if plasmid instability is mainly due to host-plasmid compatibility this suggests Ragi has evolved some means of excluding the plasmid without self-detriment, which could be a powerful cellular immunity innovation. As with the laboratory strains these two hypotheses are not mutually exclusive. It is possible that there could be a fitness tradeoff for the host even if they have evolved a plasmid defense mechanism.

3.6 Discussion: Plasmid loss dynamics consistent with missegregation mechanism

Understanding the pattern of plasmid loss in a host population might suggest possible molecular mechanisms underlying this phenotype. Because the flow cytometry assay measures single cells, and GFP expression levels correlate well with DNA copy number in budding yeast, I expect that population distribution of plasmid abundance would reveal underlying loss mechanism dynamics (Chapter 2). For example, if the plasmid were undergoing systematic underreplication I would observe an overall decrease in median plasmid copy number across the population. Alternatively, if the plasmid were being missegregated I might instead see population heterogeneity, with some cells inheriting no plasmid, and others maintaining or receiving an increased number of plasmids.

My data show that the majority of plasmid loss events are due to a complete loss of plasmid, rather than an overall decrease in copy number (Fig 3.5). This quantitatively mirrors what I

qualitatively observe by colony sectoring (Fig 3.2.a). Ragi cells that have the plasmid remain on average as GFP-intense as those under selection; however, the proportion of cells with no GFP (no plasmid) increases significantly. I suspect that there is not a substantial increase in super-green cells (extremely large plasmid burden), as others have published that cells with too many 2-micron plasmids fail to divide, presumably due to replication stress. This observed pattern of plasmid loss is more indicative of plasmid segregation failure during host cell division, rather than a copy number suppression mechanism or plasmid underreplication.

These results are especially tantalizing since the mechanism of 2-micron plasmid segregation has been proposed, but has not yet been conclusively proven[36]. In addition, we currently lack a complete picture of which host components are involved in this process, although the Jayaram lab and others have made significant contributions to this field. Perhaps understanding how some host cells interfere with this process may reveal further mechanism information about how plasmid segregation takes place in permissive strains as well. In this way using natural variation and evolution may facilitate molecular insights otherwise challenging to parse through traditional genetic and biochemical methods.

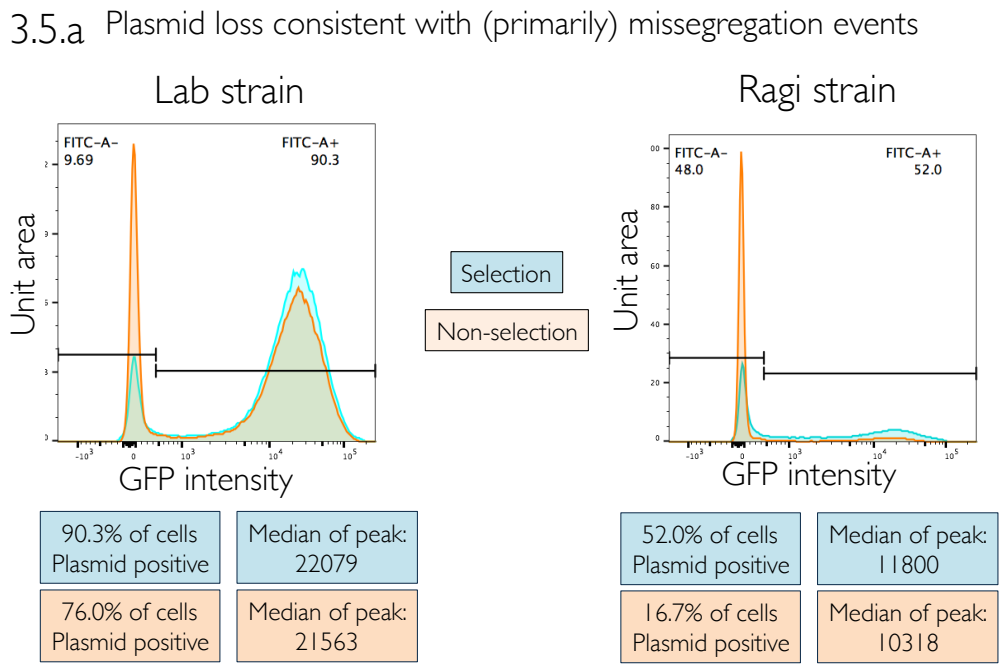

Figure 3.4: Plasmid loss consistent with missegregation as primary mechanism of plasmid loss, although median GFP positive peak intensity reveals stably lower plasmid copy number in GFP positive cells.

3.7 NATURAL VARIATION IN PLASMID STATUS AND SEQUENCE

Others had reported variation in plasmid presence before, and after the outset of this study other groups have also looked at natural variation in plasmid abundance[33,93,96]. However, ours is the first study to test if this variation is due to heritable or stochastic loss. My plasmid status findings are consistent with these subsequent publications: the isolates I tested above are plasmid-negative in others' hands as well, as assessed by whole genome sequencing. Additionally, other strains published to be plasmid negative in these other collections were plasmid containing in my hands. This discrepancy could be due to stochastic plasmid loss or sequencing library preparation done by others, or because of PCR contamination in this study.

In either case, in this work I was not overly concerned about the possibility of false plasmid-containing strains, rather I was focused on identifying truly plasmid-negative strains.

These more recent studies open up exciting future directions that were not possible at the outset of this study. Now with more known about natural variation of the plasmid sequence within *S. cerevisiae*, as well as a larger collection of host genetic variation and plasmid occupancy data, broader questions about host-plasmid compatibility can be explored. I discuss this further in Chapter 5, but as one example it would be exciting to pair SCAMPR with different plasmid backgrounds and explore host range. Are natural isolates that are resistant to one plasmid susceptible to other variants? Conversely, perhaps some hosts are broad restrictors of plasmids, while others have more targeted "immunity" to a specific subtype.

3.8 METHODS

3.8.1 *Screening for endogenous 2-micron plasmid*

Natural isolates (see strain table in Table 1.1) were generously shared by Dr. Justin Fay. DNA from these strains was isolated using standard Hoffman and Winston preps, then probed by PCR and Southern blot[89]. Briefly, primers were designed to match REP1 and FLP - the two best conserved coding regions of the plasmid as described at the time of this work (circa 2014)[97]. Only strains that showed negative PCR results for both sets of primers were further validated by chemiluminescent Southern blot using the Thermo North2South kit (17097) per manufacturers instructions. Briefly, whole genome DNA was sheared, run on an agarose gel in TAE, DNA was transferred to membrane and was probed with chemiluminescent probes created from digested endogenous 2-micron plasmid collected from BY4741 by Zymoresearch yeast plasmid miniprep kit, per manufacturers instructions. Southern blot conditions should have

permitted probing of polymorphic sequence in the event a diverged plasmid was present but undetected by PCR.

3.8.2 *Colony sectoring*

Colony sectoring was recorded by plating cells grown with G418 selection in liquid media onto YPD nonselective plates. Colonies were imaged after 2 days growth at 30C on a Leica dissection scope (for flies, not yeast spores) with a GFP filter. Image processing to split channels and color the GFP channel was done in ImageJ. No further image manipulation was done other than to trim field of view for figures.

3.8.3 *Plasmid loss assays*

SCAMPR and MCM assays were done as previously described in Chapter 2.

Chapter 4. CHAPTER 4: MAPPING THE GENETIC DETERMINANT(S) OF RAPID 2-MICRON LOSS

4.1 INTRODUCTION

The 2-micron is a putative selfish element; this plasmid relies on the host for survival, but as I and others have measured, hurts host fitness in the process. I wondered if natural variation in host genomes and plasmid resistance might reveal more about plasmid biology, host evolved immunity, and the interfaces that host and parasite are in conflict. I found that 2-micron resistance in *S. cerevisiae* is heritable and dominant, but seemingly rare within the population. I decided to focus on one strain to deduce the genetic basis of this resistance.

Here I describe my genetic mapping approach, the identified genomic loci, and efforts to further fine map the causal variants within the largest contributing locus. I chose to use a genetic mapping strategy comparing one of the wild isolates I characterized in Chapter 3, to the BY4742 laboratory strain to discover what natural variants give rise to this trait. For simplicity I selected Ragi (Y9; see Chapter 3) to perform QTL mapping for three reasons. First, it had the least variable plasmid loss phenotype, which was most distinct from BY (the haploid parent permissive of plasmid maintenance). Secondly, others had published that in their hands the NJ oak strain, YPS1009, was aneuploid[98]. I thought this might complicate genetic mapping, so opted to avoid this specific strain for the cross. The Ragi diploid sporulated well, so I speculated this strain may be euploid. Finally, others have seen Y9 and Y12 are closely related and both cluster with the mosaic sake strains[26,30]. As such, I speculated that these 2 strains may in fact share the same haplotype for the region(s) of the genome responsible for 2-micron restriction[32]. This close relation could, in conjunction with additional isolates for comparative genomics,

provide a useful tool for narrowing down which polymorphisms in a genomic region might be causative for plasmid loss.

4.2 RESULTS: PLASMID LOSS IS A NON-MONOGENIC TRAIT

Is rapid plasmid loss a monogenic or complex trait? Mapping complex and simple traits can require different experimental design and methods, so I first wanted to explore how many unlinked loci are likely to drive the bulk of this trait. To better understand the genetic architecture underlying this rapid plasmid loss trait I looked at progeny phenotypes from a cross between BY4742 and a stably haploid Ragi strain. If a single genetic locus is responsible for this trait I would expect all tetrads to exhibit a 2:2 segregation pattern, with half of the spores phenotypically resembling the BY parent and the other half resembling the Ragi parent. If the trait is polygenic the inheritance pattern will be complex, with a greater number of independently segregating loci corresponding to an increase in the number of progeny with intermediate plasmid-loss phenotypes. My results indicate that this is not a monogenetic trait (Fig 4.1.a), with only about 20% of 4 spore tetrads exhibiting a roughly 2:2 segregation pattern, and other tetrads demonstrating more complex patterns or inheritance. Because plasmid loss is a dynamic trait (a loss rate rather than a binary phenotypic state) these data are somewhat noisy. To account for this phenotypic variation, all progeny were measured in triplicate, and error bars reflect the standard deviation of replicates for plasmid loss measured by flow cytometry. Even accounting for this inherent noise, it is obvious that not all tetrads exhibit a segregation ratio consistent with a parental ditype. Instead, many tetrads show plasmid loss patterns indicative of more complex trait architecture. These data demonstrate that more than one genomic locus is likely responsible for the variation in plasmid stability between the parental strains; however, with ~20% tetrads showing approximately 2:2 plasmid loss ratios, I infer that there are only a few large effect loci segregating for this trait. This genetic architecture (although not monogenic, still

not very complex) is still addressable by bulk segregant analysis approaches, where a more complex trait might not be.

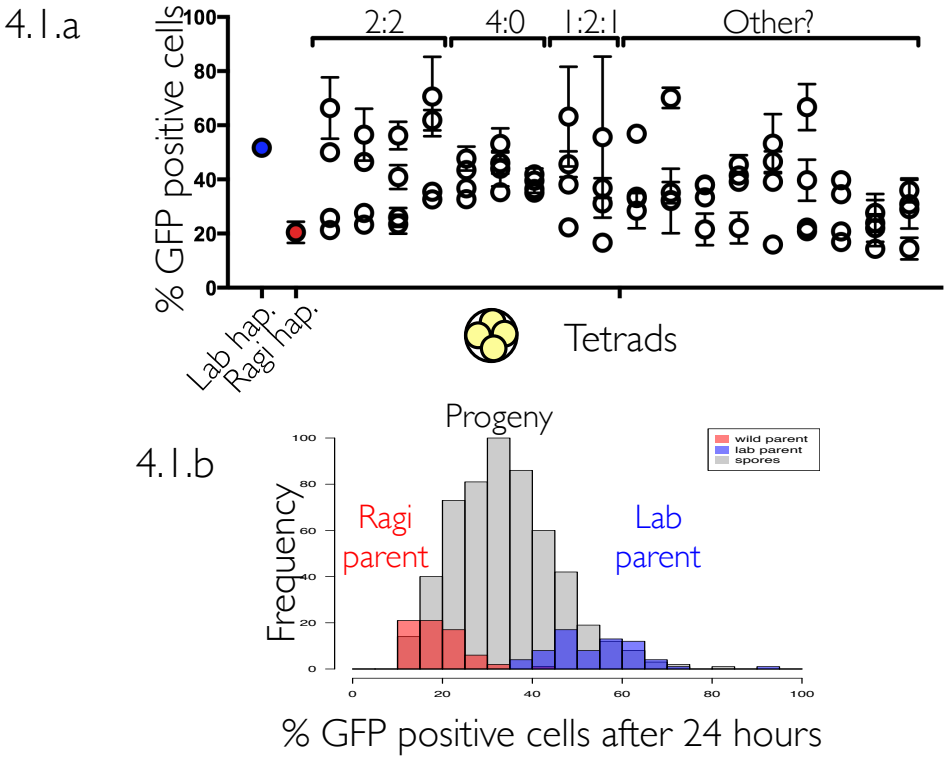

Figure 4.1: Analysis of progeny phenotypes. 4.1.a) Tetrad analysis reveals non-monogenic trait. 4.1.b) distribution of progeny phenotypes relative to parents across experiments

The distribution of progeny phenotypes (here plotting the mean of 3 replicates per progeny strain) can also be compared to the variance in parent frequencies (compiled across experiments) (Fig 4.1.b). In this graph the sister spore relationships are lost, but one can more easily visualize the distribution of progeny phenotypes from a larger cross.

4.3 GENETIC MAPPING EXPERIMENTAL DESIGN

Because the plasmid loss trait in Ragi seems to be multigenic, but not very complex, I opted to employ quantitative trait locus (QTL) mapping to resolve which genomic loci are tied to rapid plasmid loss (Fig 4.2). This approach is powerful because it can determine genomic regions

associated with a trait of interest without prior knowledge of the underlying genetic architecture. Indeed, QTL mapping can be used with recessive or dominant traits, and monogenic or complex traits. This approach identifies loci linked to a phenotype of interest by looking for correlation of one parent's haplotype with a trait of interest in a pool of progeny.

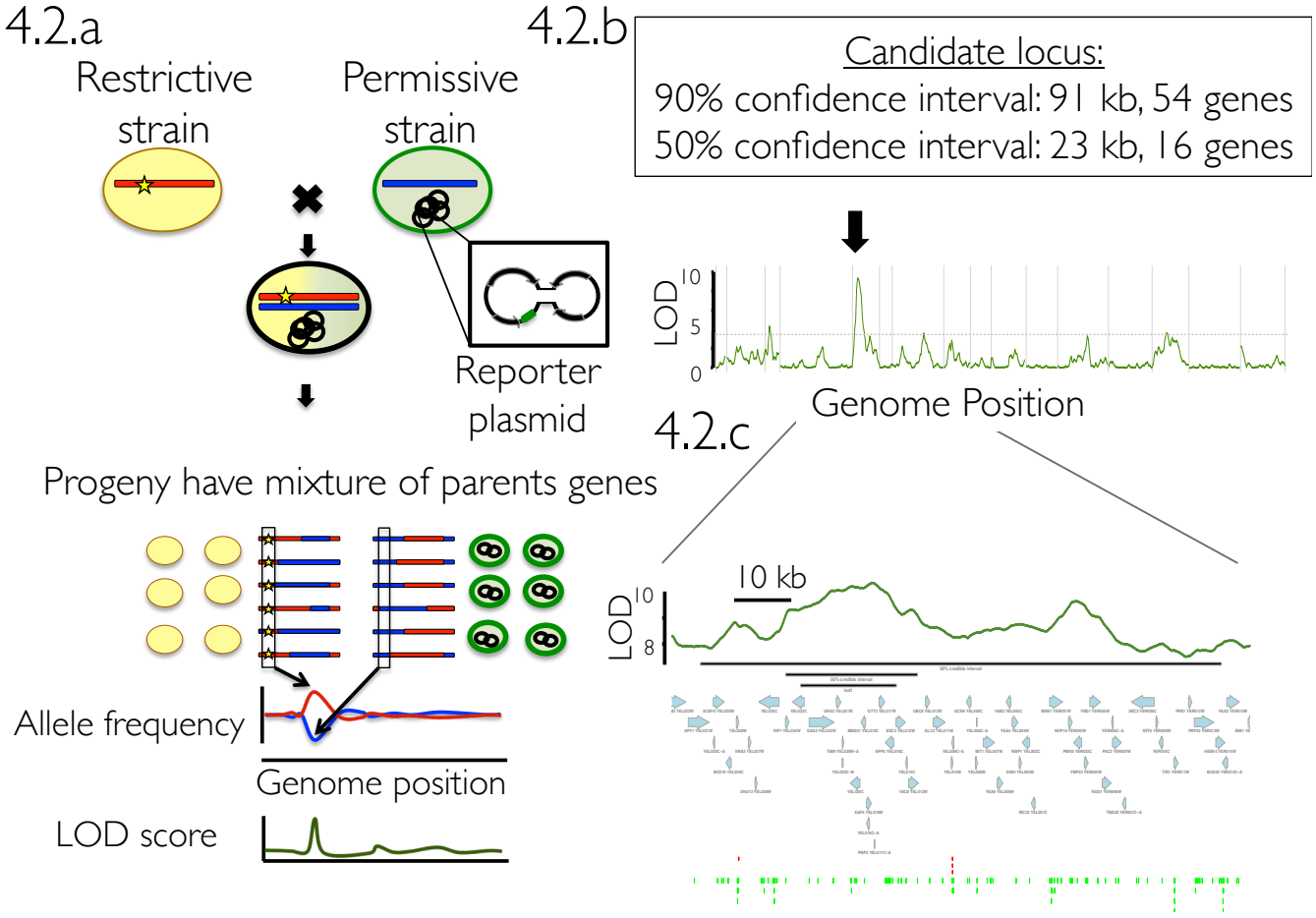

Figure 4.2: QTL mapping and results 4.2.a) experimental design. 4.2.b) QTL mapping identified a locus of interest on Chromosome V associated with rapid plasmid loss. Panel b depicts whole genome LOD scores. 4.2.c) Inset is a close-up of Chromosome V peak.

Because my previous analysis suggested that this trait has only a few large effect loci contributing to loss, I chose to do mapping by bulk segregant analysis (BSA) rather than

sequencing individual progeny. Progeny are pooled based on phenotype, then the pooled are whole genome sequenced. Statistically significant deviations from the expected random allele frequency of ~0.5 are associated with regions of the genome responsible for, or genetically linked to, the phenotype being mapped. Although this approach loses the ability to associate specific haplotypes to specific progeny, the bulk approach is cost effective because so many fewer sequencing libraries need to be prepared. Briefly, I pooled progeny by plasmid phenotype as determined by flow cytometry ('stably inherited' vs. 'rapidly lost'), then whole genome sequenced each pool and looked for regions linked to plasmid restriction by allele frequency. After the initial pilot experiment investigating segregation patterns the following genetic mapping studies were done by random spore analysis. To enrich for parental allele combinations in the sequenced samples, I pooled the 20% of progeny with the highest GFP stability and the 20% of progeny with the lowest GFP stability.

Although rapid plasmid loss is a dominant acting trait (Chapter3) in the heterozygous (Ragi x BY4742) diploid, I elected to introduce the plasmid into the parents of the cross and into the progeny through meiosis rather than by transformation of the progeny themselves. For practical reasons this approach was more straightforward. The reporter 2-micron plasmid does not have bacterial cloning sequence by design, in order to minimize the changes to the endogenous SGE, and to avoid destabilizing the plasmid as others have measured. However, recovering plasmids from yeast DNA preparations is highly inefficient. This inefficiency made getting enough plasmid to do large-scale transformations (hundreds of progeny strains) technically limiting. However, I was able to ensure with a plasmid selectable marker, that only those progeny that received the 2-micron during meiosis were used. Additionally, I knew from previous tetrad analysis (Chapter 3) that I was able to recover 4-spore tetrads in most cases. This finding means the putative restriction factor is not so fast acting that I would be unable to recover

progeny with the necessary genotype to perform mapping, so long as strains were maintained on selection and stored at -80C promptly.

While QTL mapping and BSA are powerful because of the flexibility to map traits of differing or unknown genetic architectures, one limitation is that often the resulting loci identified are quite large, and encompass multiple genes and polymorphisms. Fine mapping which variants within the loci are actually responsible for the trait of interest can be done in multiple ways, but many are labor and time intensive. Additionally, these methods often do require knowledge of (or assumptions about) underlying genetic architecture. Below I discuss the fine mapping strategies employed here, as well as my justifications for those methods, and alternative strategies.

4.4 RAGI PARENT STRAIN IS DISOMIC FOR CHROMOSOME XIV

I had initially chosen Ragi for QTL mapping over the North American oak strain (YPS1009) because it was known to be aneuploid in other's hands. However, aneuploidy is somewhat common in *S. cerevisiae* isolates. The Ragi diploid I began with sporulated well, and gave rise to viable progeny so I did not initially suspect that the parent might be of unusual karyotype. In fact, whole genome sequencing revealed that my original Ragi parent haploid strain was disomic for chromosome XIV (Fig 4.3). Additionally, it was unclear at the time if my parent diploid was aneuploid, or if I had selected an unfortunate aneuploid novel haploid isolate during the process of making homothallic stable haploids for this project. This aneuploid chromosome segregated in the Ragi x BY cross. I wondered if this additional chromosome may influence plasmid stability. To test this possibility I identified a euploid Ragi strain for chromosome XIV by qPCR probes (previously designed by Rong Li lab), and then measured plasmid loss in the euploid background (Fig 4.3)[99]. Regardless, this independently isolated and presumed euploid Ragi strain showed the same plasmid loss phenotype as the original isolate, suggesting that the

aneuploid chromosome is not a large contributor to my phenotype of interest, and as such the segregating disomy was largely ignored in my subsequent experiments.

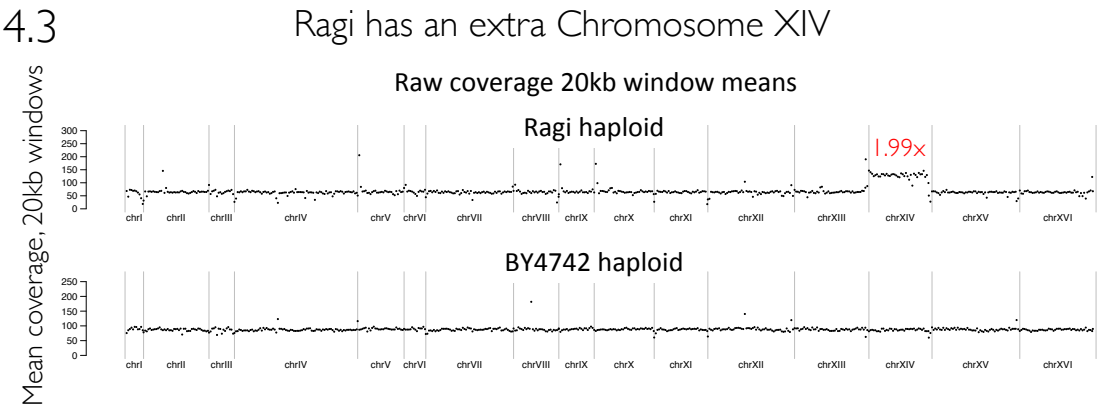

Figure 4.3: Ragi parent is aneuploid for Chromosome XIV

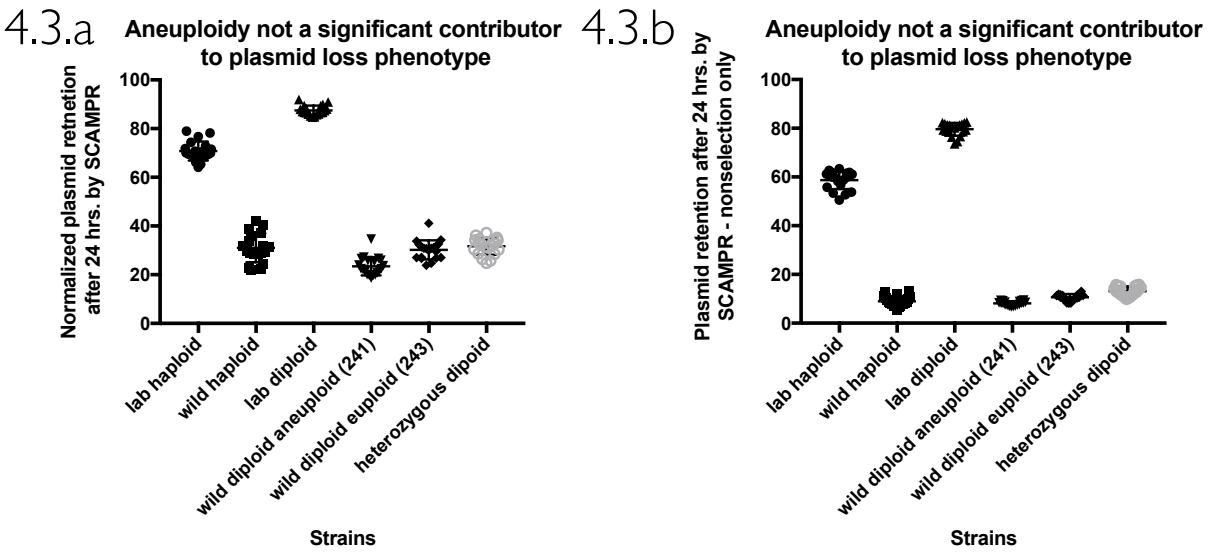

Figure 4.3: Euploid Ragi shows similar plasmid loss as Chr XIV aneuploid. Here I show both the normalized (4.3.a - corrected for starting values under selection) and non-selection only condition (4.3.b)

4.5 GENETIC MAPPING OF PLASMID LOSS BY BULK SEGREGANT ANALYSIS

To identify what genomic differences (and eventually what molecular mechanisms) underlie this plasmid loss I took an unbiased quantitative trait locus (QTL) mapping strategy (Fig 4.2.a). Because I know from tetrad analysis that the plasmid loss trait is multigenic, but is not very complex (probably less than 3 large effect loci) I employed bulk segregant analysis to identify which loci contribute to rapid plasmid loss in one of the natural isolates, Ragi. I chose to focus here on a single isolate, but ultimately am interested in understanding if all three isolates share underlying mechanisms (or indeed alleles) or if different strains have avoided plasmid parasitism through different means. My results highlighted one large confidence QTL with a LOD score of ~10 on chromosome 5, with a couple of peaks with LOD scores closer to 4 (Fig 4.2.b). The locus corresponding to the highest LOD score covered ~91 kb (54 ORFs) for the 90% confidence interval (as determined using Multipool), and 50% confidence interval covered 23 kb (16 ORFs)[100]. I chose to focus my fine mapping efforts on this Chromosome V locus because of the large LOD score; however, I assume that other QTL are also contributing to this trait based on the tetrad segregation pattern.

While the Chromosome V locus is very compelling and well supported, that 90 kb still included several polymorphisms between the parental genomes that could be responsible for this trait. Relative to BY, Ragi has 47,173 SNPs across the entire genome. Two frameshift and 2 nonsense SNPs lie within the Ragi 90% CI, all of which fall in dubious ORFs; 107 missense mutations fall within 33 Ragi genes.

4.6 FINE MAPPING: THREE STRATEGIES AND ASSOCIATED RESULTS

To explore which of these variants might be causative for the plasmid loss trait I decided on a three-pronged strategy for fine mapping (Fig 4.4). 1) Some candidate genes were tested for sufficiency by inserting a copy of the Ragi allele into a BY4742 haploid. Candidate genes were chosen for reasons outlined below, but comparative genomics helped prioritize which SNPs were most likely to be associated with plasmid loss. 2) Candidate genes and smaller subregions of the locus (~5-10 kb) were be tested for necessity by knocking out the Ragi allele (creating hemizygosity for the gene of interest) in the heterozygous diploid. 3) Loss of heterozygosity was induced by mitotic recombination across the locus to narrow down the region of interest.

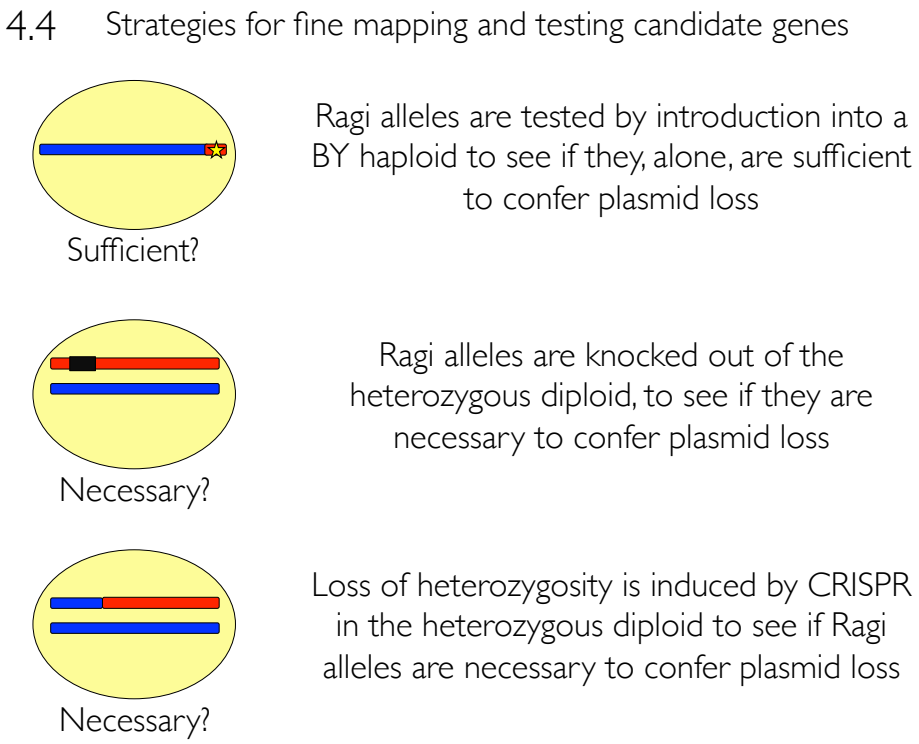

Figure 4.4: Fine mapping strategies chosen for this work

4.7 CANDIDATE GENES

Based on what is known about plasmid function within the host I chose candidate genes within the chromosome V locus that had protein coding differences between BY and Ragi and

compelling circumstantial evidence (detailed below) that they might be affecting plasmid stability. I further prioritized SNPs by looking at which were shared between Y9 and Y12 (closely related strains that are both restrictive of plasmid) and are not shared with other natural isolates that do harbor endogenous plasmids such as PW5.

I tested the following candidate genes:

YEL025c is a gene of unknown function that lies within the 90% confidence interval. It has a single nonsynonymous polymorphism between Y9/Y12 and out groups. I found this candidate gene exciting because it was previously seen to be rapidly evolving when compared between *S. cerevisiae* and *S. paradoxus* by sliding window dN/dS[101]. In addition the protein has been seen to localize to the nucleus, which would be required to directly interact with the 2-micron plasmid, as well as the cytoplasm[102].

MMS21 is a mitotic SUMO E3 ligase (one of 3 in *S. cerevisiae*)[103]. The Mms21 protein is a member of the SMC5/6 complex, a circular protein complex important for replication and repair of high copy number DNA which lies at the nuclear periphery. It anchors dsDNA breaks and facilitates resolution of complex DNA structures[104]. Because we know the 2-micron requires sumoylation of its protein products for proper function and it also physically locates to the nuclear periphery (and is high copy number and may result in complex DNA structures if FLP1 is activated) I suspected that this polymorphism might influence plasmid stability within the host cell[73].

Finally, *URA3* also falls within the 90%CI locus; this gene was especially concerning to me because BY4742 is an uracil auxotroph, and this gene often falls within QTL related to fitness in BY crosses[105,106]. I became worried that, although all experiments were performed in rich media, that somehow the *ura3* deletion was leading to the BSA signal. To confirm that *URA3* status is

not responsible for this measured difference in plasmid stability between strains, I also knocked out *URA3* in Ragi and saw no increase in plasmid stability (Fig 4.5). As further support I know that the 2-micron plasmid is stable in many other prototrophic lineages, with the majority of natural isolates, as well as prototrophic laboratory strains, found harboring endogenous plasmid. Taken together these lines of evidence led me to rule out URA3 as a causative locus for differential plasmid stability between these 2 strains.

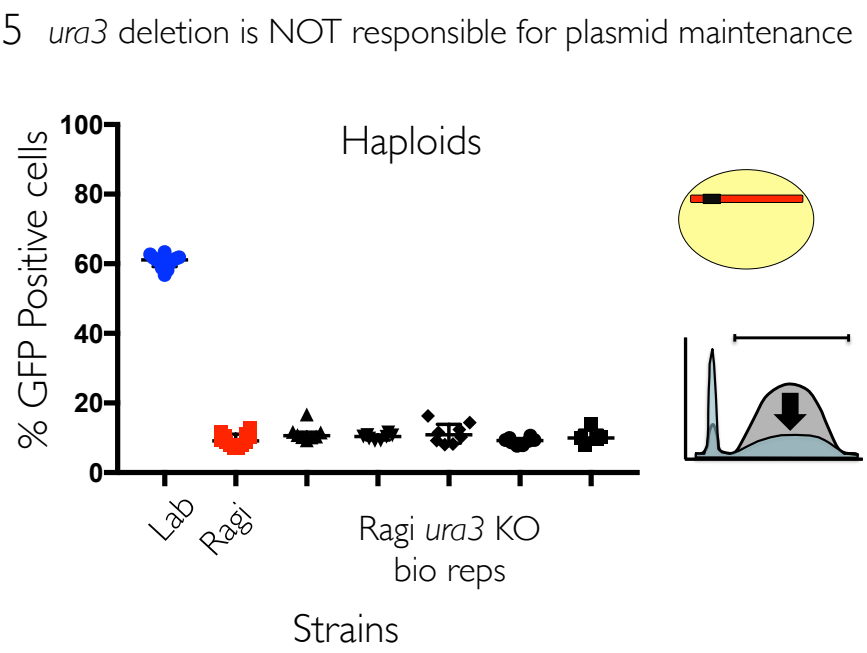

Figure 4.5: *URA3* not responsible for plasmid maintenance

4.7.1 *Sufficiency testing*

I wondered if either Ragi *MMS21* or Ragi *YEL025c* alleles was sufficient to confer the plasmid loss trait. Because I previously determined plasmid loss was a dominant trait, I tested if the polymorphic allele of either gene alone was sufficient to induce plasmid loss in the BY haploid (Fig 4.6). I integrated the Ragi allele, with the flanking intragenic regions, into the *ho* locus of BY4742 and measured plasmid stability (Fig 4.6). My results show that neither Ragi *MMS21* or

YEL025c is sufficient in BY4742 alone to confer the plasmid loss trait. I note that the addition of the *MMS21* Ragi allele to BY does slightly reduce plasmid stability, but the difference is not statistically significant as determined by a Kruskal-Wallis test, corrected for multiple comparisons. This is consistent with the possibility that the *MMS21* variant is partially responsible for the plasmid loss trait, but requires other loci to achieve full plasmid loss.

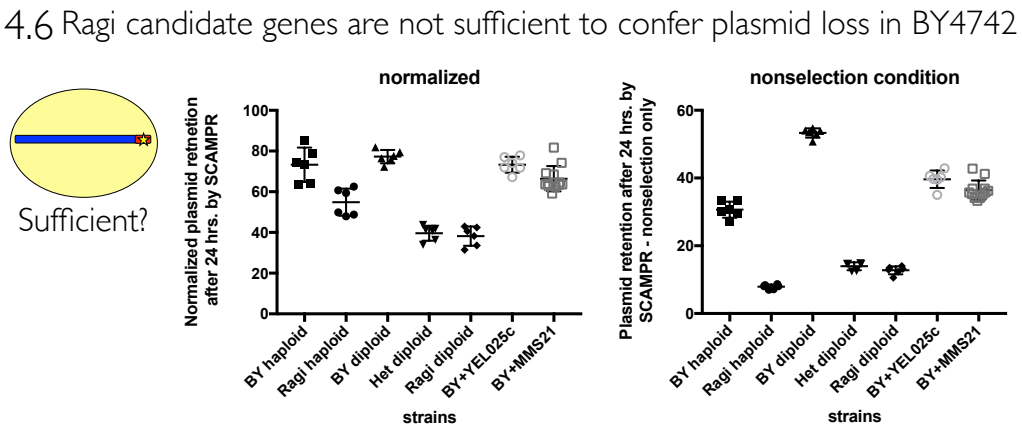

Figure 4.6: Ragi candidate alleles not sufficient for plasmid loss in BY4742

4.7.2 *Necessity testing*

However, because I had previously seen that this trait appears to be multigenic, and I do not know the genetic architecture underlying this trait, it is possible that these alleles are necessary (though not sufficient) to confer plasmid loss. To test this possibility I sought to knock these alleles out and see if plasmid resistance was lost. Because *MMS21* is an essential gene (as well as others in this genomic region), I did this in the heterozygous diploid background so that the strains are kept alive via the BY alleles. *YEL025c* was not necessary for rapid plasmid loss in the heterozygous diploid background (Fig 4.7), however the *mms21* hemizygote showed a small, but reproducible and statistically significant increase in plasmid stability (Fig 4.8). Furthermore this gain in plasmid stability was specific to a mutant missing only the Ragi allele of *MMS21*. The same heterozygous strain missing the BY allele of *MMS21* did not show a similar increase in plasmid stability. These results suggest that while MMS21 is one contributing

polymorphism in the Chromosome V peak, it is not the only allele underlying rapid plasmid loss. This is unsurprising given I already knew it was a multigenic trait based on tetrad segregation. However, it also suggests that the Chromosome V BSA signal is most likely due to multiple linked polymorphisms, as opposed to one large effect variant. Because of the unknown genetic architecture, as well as technical difficulties in genetically manipulating Ragi, we chose not to pursue the other contributing genetic variants within this region at this time.

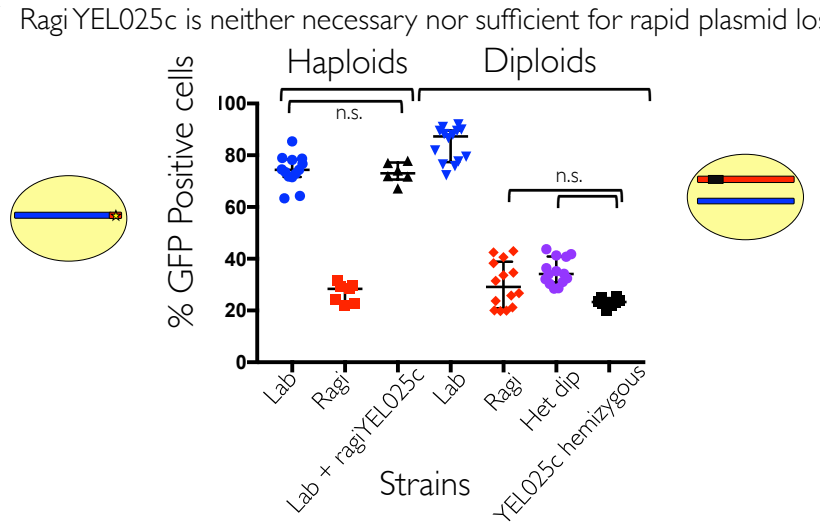

Figure 4.7: Ragi *YEL025c* allele not sufficient to confer rapid plasmid loss in BY4742, nor necessary in heterozygous diploid background

4.8 Ragi *MMS21* underlies a small, but significant, portion of rapid plasmid loss phenotype

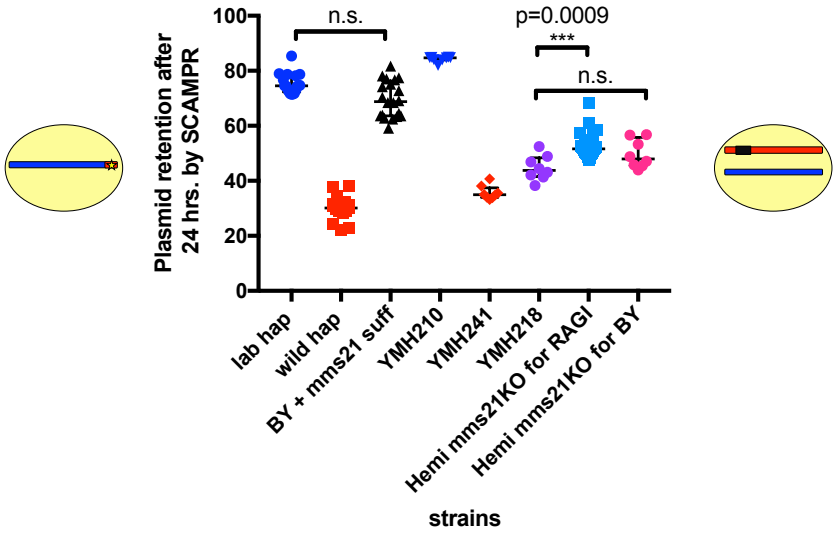

Figure 4.8: The Ragi *MMS21* allele is not sufficient alone to confer rapid plasmid loss in BY4742. However, Ragi *MMS21* shows small but statistically significant contribution to plasmid loss in the heterozygous diploid. This is dosage independent as the BY allele does not confer the same degree of plasmid loss.

In addition to taking a candidate gene approach, I also began undertaking two more agnostic fine mapping strategies; first to expand the hemizygosity testing to further subregions within the 90%CI peak (Fig 4.9), and second to engineer loss of heterozygosity (Fig 4.10) via a CRISPR-induced mitotic recombination strategy. The “ura left” region knocked out ~14 kb of Ragi alleles spanning *YEL025c* to *URA3*, while the “ura right” region knocked out ~5 kb of Ragi alleles between *MMS21* and *URA3*. To date these approaches have not yielded any further candidate polymorphisms, although these mapping efforts are ongoing within the lab. Several alleles have been knocked out in the hemizygosity experiments spanning approximately 25 kb of the locus of interest. All phenotyped mutants to date have not shown statistically significant changes in plasmid stability relative to the wild type heterozygous diploid. I suspect that there may be multiple linked variants in the Chromosome V peak. Because these knockouts were not created

all at the same time in the same strain (due to marker availability) I have not been able to test for additive or synergistic plasmid loss of these variants. Currently my MMS21 results and subsequent hemizygous phenotyping leads me to suspect that there may be multiple linked contributing variants within this Chromosome V locus, in addition to other independently segregating loci that contributed to the observed tetrad segregation pattern.

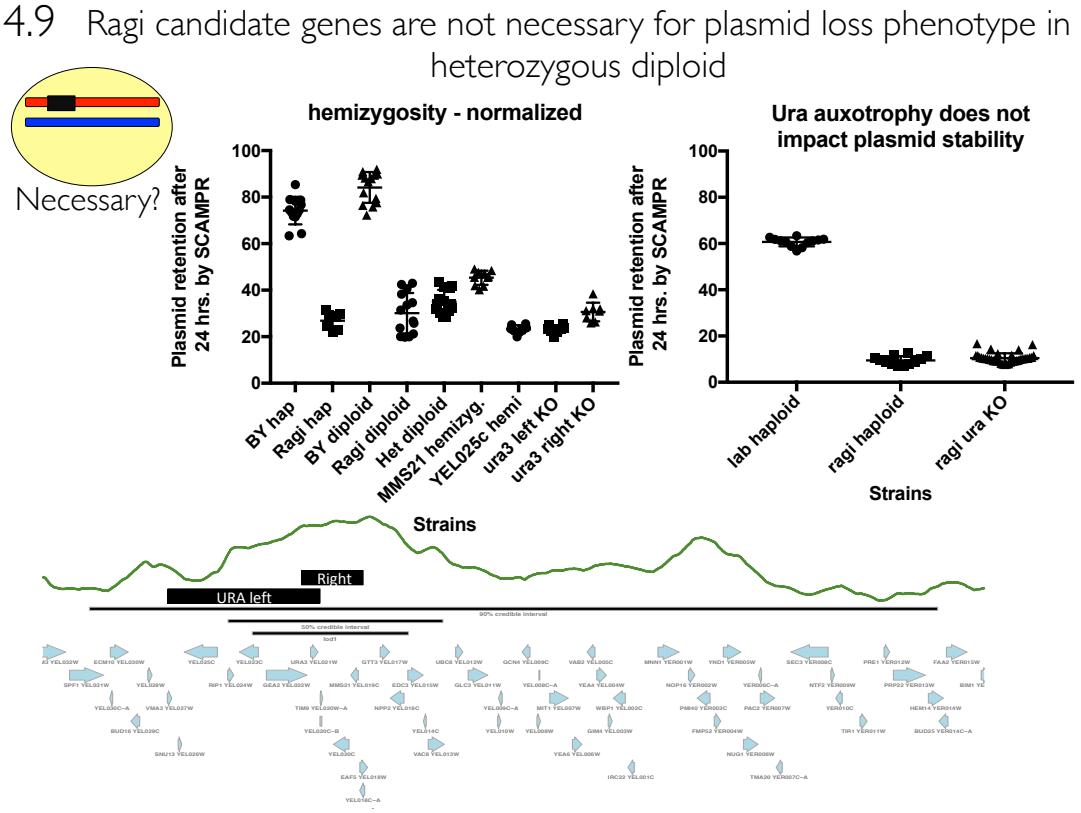

Figure 4.9: Knock-outs of Ragi alleles in the heterozygous diploid does not significantly improve plasmid stability. Deletion of *URA3* in the Ragi haploid does not confer plasmid stability. Although the Ura-right knock out includes *mms21*, in this experiment the difference in plasmid stability was not statistically significant, however the trend towards more stability holds true. This could reflect experimental variation, or a complex genetic architecture for this trait.

4.8 Restriction or Haploinsufficient permissivity?

Because of my prior experimental data showing that the heterozygous diploid rapidly loses plasmid I believed that rapid plasmid loss is most likely a dominant trait. However, it is formally

possible that the trait is haploinsufficient instead. I initially believed haploinsufficiency would be unlikely, however, because it would require the host encoded gene be completely dispensable for host function (Ragi haploid strains are viable and grow well), but be essential AND haploinsufficient for plasmid function. This scenario would be surprising because the host encoded permissivity factors that have been previously described fulfill the same function for both the host and plasmid, such as replication origin licensing machinery and sumoylation factors. A factor that is essential for the plasmid, but does the same job for the host but is non-essential for host function would be surprising. However, this situation is formally possible. One could imagine a scenario where a host factor serves different functions for the host and plasmid, for example. Or perhaps the host has redundancy (e.g. another protein can replace the host function in a mutant background) while the plasmid is not able to tolerate this substitution. As such, a mapping strategy that is effective in identifying both haploinsufficient recessive alleles and truly dominant alleles equally would be ideal. Previously Sadhu *et al*. described a method for inducing mitotic recombination in lab strains of *S. cerevisiae*. I sought to implement the same conceptual strategy, but required tools that would function in the Ragi/reporter 2-micron strain background[107].

4.9 Loss of Heterozygosity (LOH)

Accordingly, I, with the help of Paula Levan, built the following reagents to facilitate CRISPR driven LOH in the Ragi x BY heterozygote (Fig 4.10). As with the original method, double stranded breaks on a single chromosome were introduced by CRISPR, driving homologous repair with the intact chromosome as the repair template. Guide RNAs were designed to specifically target PAM sites unique to only the Ragi Chromosome V, to break up the locus of interest while maintaining stoichiometry with the rest of the genome. In this way this method facilitates mapping of either a truly dominant trait, or a haploinsufficient recessive one.

Because Ragi does not maintain 2-micron plasmids well, I was concerned about introducing 2-micron-based episomal vectors based on the 2-micron as a means to create genetic modifications to test my trait. The original CRISPR LOH method utilized both CEN and 2-micron based vectors to induce gRNAs and Cas9 respectively. To avoid possibly confounding the plasmid loss trait with strain construction efforts, I built a single CEN vector that contains both the Cas9 gene and gRNAs, designed with a drug resistance marker to work in prototrophic strain backgrounds. To select cells in which the chromosome arm of interest had undergone an LOH event, I integrated in dsRed fluorescent markers on the distal arm of Chromosome V. Strains that successfully repaired off the BY chromosome would no longer fluoresce red, and can be sorted by FACS.

While these molecular tools have been created, the efforts at fine mapping by CRISPR LOH are still currently underway.

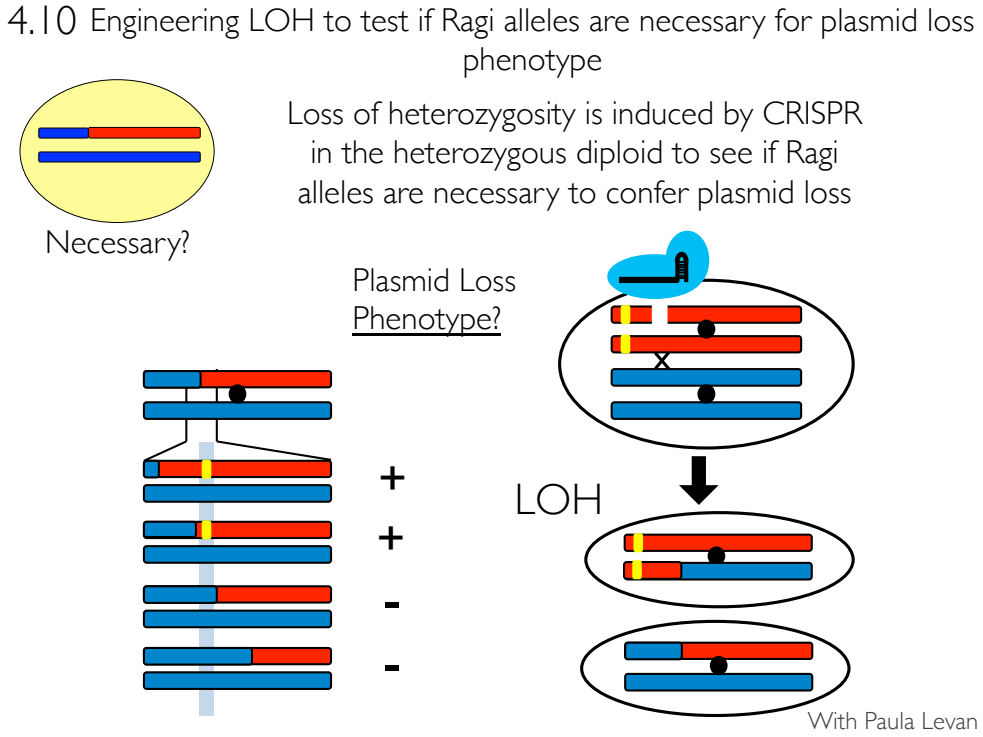

Figure 4.6: Engineering LOH within the Chromosome V locus via CRISPR-direct dsDNA breaks

4.10 DISCUSSION AND FUTURE DIRECTIONS

Here I sought to map the causal variants for a naturally evolved host resistance to the 2-micron plasmid. I hoped that a QTL and fine mapping approach would reveal novel molecular insight into how host and plasmid interact. Although this work is still ongoing, we've identified a small, but compelling genomic locus associated with this trait. It will be exciting if this work leads to the discovery of a bona fide host factor that dictates plasmid susceptibility. Leveraging the natural variants that are present in the *S. cerevisiae* population today will help us understand how host-plasmid coevolution is shaping these genomes.

Because I had observed variation in *S. cerevisiae* susceptibility to 2-micron plasmid, I wanted to determine what the causal genetic variants were that give rise to this resistance. I selected one natural isolate to perform QTL mapping, and identified one locus with a compelling LOD score of ~10. I employed multiple strategies to determine what within the region is the causal variant, but as of yet have only ruled out polymorphisms and an unexpected aneuploidy as neither necessary nor sufficient, although some avenues remain open for identifying the underlying causal variant.

While a number of ORFs and subregions of the primary chromosome V locus have been eliminated as neither necessary nor sufficient for rapid plasmid loss alone, and a single variant has a weak but reproducible impact on plasmid stability (*MMS21*), a complete understanding of the causal variants remains elusive. This could be due to a number of reasons. I suspect the underlying genetic architecture of this trait makes fine mapping difficult: it is possible the QTL identified actually harbors a number of linked variants, all of which are contributing to the phenotype, and while they segregate together typically in a cross, further fine mapping is difficult. This is supported by my *MMS21* hemizygosity results. Or perhaps the trait is mostly

haploinsufficient, rather than dominant. While I considered this later possibility less likely, given that host functions remain unimpaired for growth in the haploid Ragi isolate, it is formally possible. It would require that the allele present is haplosufficient for host function, while haploinsufficient for plasmid function. Both possibilities are interesting for understanding the possible mechanisms by which a host genome can evolve to "fight back" against a genetic parasite. The LOH method outlined above should work on either haploinsufficient or dominant traits; however, tool construction challenges in Ragi has led to slow experimental progress. Nevertheless, future work exploring the genetic and mechanistic basis for this natural variation in parasite resistance will be exciting.

If the trait is not truly dominant, and is instead haploinsufficient, that raises interesting ideas about the underlying mechanisms of the trait. While a dominant trait is suggestive of a restriction factor, haploinsufficiency would suggest a permissivity factor. However, if there is a host permissivity factor that is not only essential but haploinsufficient for the 2-micron, but at the same time is completely dispensable to the host genome it raises the question: why haven't all yeast strains gotten rid of this susceptibility factor? It is possible that there are cost-benefit tradeoffs to the host in the presence of this factor. One could imagine a scenario where a host cell has some redundancy to protein function, and under laboratory growth conditions missing factor "A" would not create a large fitness cost because a different protein (factor "B") could cover its function. If factor A could not be substituted for maintaining plasmid function however, the host may be able to lose factor A to restrict the 2-micron plasmid. Perhaps there are other conditions (such as environmental stress) in which factor A and B are not interchangeable for host function. Knowing what those tradeoffs are for the host, may help inform more about basic plasmid biology as well as what the interfaces of coevolution between the two genomes are.

These same cost-benefit tradeoffs may also impact the prevalence of this plasmid loss trait. This may, in part, may explain why the plasmid is it so prevalent within *S. cerevisiae* even though it confers a fitness cost to the host. Although there is currently no evidence to support it, it is possible that there are in fact cryptic benefits to the host for having the plasmid. Perhaps I have not seen these benefits in laboratory conditions. Or perhaps the fitness costs associated with heritable plasmid restriction are great enough that there is little benefit to these polymorphisms in the wild. Knowing what the underlying genetic variants are, and eventually what the molecular mechanism is, will allow us to differentiate these possibilities.

Through the lens of *MMS21* natural variation it will be interesting to explore host constraints and fitness tradeoffs for plasmid-resistant host alleles. Is this MMS21 variant a hypomorphic allele: able to support host viability without the burden of the 2-micron plasmid, but suffering a synthetic fitness cost when the 2-micron is present? Alternatively it is possible that this polymorphic allele is actually a separation- or gain-of-function mutant: somehow able to fully perform host functions while excluding the plasmid. To test between these possibilities it will be exciting to have an expansion of the SCAMPR assay that allows for measuring fitness cost of the plasmid in different genetic backgrounds while monitoring plasmid loss simultaneously (Fig 2.10). Possible fitness tradeoffs are high, given MMS21 is essential for the host. Structure and function studies have determined that the MMS21 protein domain that the Ragi polymorphism lies in (alpha helix three) is actually important for DNA damage response in the host, although this specific polymorphism has not been tested directly. In the future it will be interesting to see if Ragi has a fitness defect under DNA damaging conditions, although the doubling time in normal laboratory conditions is the same as the wild type lab strain. Ultimately understanding the genetics that underlies natural variation in host resistance to the 2-micron will allow us to explore why the plasmid still exists, and how is this antagonistic relationship may be driving evolution of each party.

4.11 METHODS

4.11.1 *Strain growth and construction*

Unless otherwise noted, yeast strain construction and growth was performed by standard yeast methods including lithium acetate transformation, and growth at 30C[89]. Vector construction was done in *Escherichia coli* by Gibson assembly unless otherwise noted.

4.11.2 *Illumina sequencing, library preparation*

Sequencing libraries were prepared by the TruSeq method for genomic DNA. Samples were multiplexed on an Illumina HiSeq lane and run by the Fred Hutchinson Sequencing core facility. Bulk segregant analysis pool comparison was done using the Multipool package tool for LOD score and statistical analysis with the help of Dr. Janet Young[100]. Average depth of read coverage was used to estimate karyotype and deviations from euploidy.

4.11.3 *Plasmid loss assay*

All plasmid loss assays were measured by SCAMPR as described in Chapter 2.

Chapter 5. DISCUSSION

5.1 A NATURAL AND TRACTABLE MODEL SYSTEM FOR HOST-SGE STUDIES

Genetic conflicts between host genomes and SGEs are often studied out of context. More often than not, the pathogen that caused the conflict is unavailable, so hosts are instead tested against other pathogens that most closely approximate what is thought to have driven the evolution initially (e.g., extinct viruses/transposons are replaced by extant versions or putative ancestral sequences). In these cases, reconstruction efforts are shallow, because evolution happens rapidly. In some cases, the true pathogen responsible may be unclear. For example, some antiviral host factors target multiple viruses (some potentially unknown), so the true driver of a conflict may not be clear. Finally, some conflicts require extremely artificial systems to be amenable to investigate in the laboratory. These systems can be powerful, but also are limiting in the types of biological questions that can be tackled. For example, viral minireplicon systems move portions of viral genomes onto vectors, allowing those components to be studied without creating infectious particles. This artificial system allows for serious human pathogens to be studied more easily, but only allows for the study of some questions about host-viral interaction, and additionally conflates vector and viral success in a host cell.

The 2-micron is an exciting model system for exploring host-SGE interactions because it circumvents many of the concerns above. Not only is it widely prevalent within *S. cerevisiae*, but also there are also 2-micron-like plasmids that have evolved in other yeast lineages. The 2-micron presents a rare opportunity to study a coevolved SGE; this known coevolution itself is rare amongst pathogens, with the exception of integrated intragenomic parasites (such as transposons). The challenge with transposons is their near-ubiquity in genomes: doing a traditional genetic knock out experiment, for example, has been traditionally impossible. Recent

efforts with whole-genome CRISPR/Cas9 deletions targeting endogenous retroviruses or creating genomes devoid of transposons via synthesis are only now opening some of these avenues, however they require extensive synthetic genome reengineering. As extrachromosomal DNA plasmids, 2-microns are tractable in the laboratory even by traditional techniques, as are the genomes of many of the host *S. cerevisiae* strains and even other budding yeast species as well. *S. cerevisiae* has long been one of the best studied model eukaryotes, benefitting from decades of tools and molecular insights and a collaborative research community. In more recent years, there has been an expansion of resources across the budding yeast clade to facilitate more evolution and ecology studies. Taken together, yeast and 2-micron plasmids are a uniquely powerful system to do both deep mechanistic studies as well as broad evolutionary studies.

5.2 LEVERAGING NATURAL VARIATION TO EXPLORE MOLECULAR EVOLUTION AND MECHANISM

The 2-micron plasmid was first described in the 1970s and is found widely within *S. cerevisiae*. Encoding a very small genome of its own, it is known to depend on host components for its survival. Additionally, it was seen in the early 1980s to confer a fitness cost to the cells that harbor it. Nevertheless, at the time this thesis project was started, the 2-micron had not been explored as a model for host-parasite coevolution. Budding yeasts offer many molecular and evolutionary resources to use in the lab, and the plasmid itself is genetically tractable, so this seemed an ideal model system to investigate how a host and parasite shape one another's genomes.

I elected to pursue a natural variation strategy. While plasmid polymorphism was poorly characterized at the outset of this project (only 2 variants in *S. cerevisiae* were known, and the

prevalence of either was unknown), host genomic variation was better characterized. Thanks to the efforts of a number of labs, as well as industry interest in *S. cerevisiae* as a species, numerous strains were whole genome sequenced and available to work with in the laboratory. For this reason, I decided to probe the putative host-plasmid conflict from the vantage point of the yeast genome. I tested a panel of natural isolates, selected to cover a broad range of within-species diversity, for isolates that did not contain the endogenous 2-micron plasmid. I hypothesized that, because of selective pressure from plasmid-induced fitness costs, some yeast isolates may have evolved means to heritably restrict 2-micron. This strategy can not identify a comprehensive list of all possible means of restriction, nor the full list of host permissivity factors. However, I thought that by identifying one or a few molecular means of plasmid restriction I may gain insight into the types of pathways at play, or indeed if novel immunity mechanisms had evolved. Moreover, this approach would give us insights into what resistance mechanisms had *actually* evolved rather than theoretically *could* evolve (under laboratory conditions), but that may be too unfit to propagate in the wild. I wanted to use these isolates as a means to genetically map restrictive genetic variants. To facilitate these studies I developed a new plasmid loss assay that would be rapid and scale well for large sample sets.

5.3 DEVELOPING A MORE ROBUST AND FLEXIBLE PLASMID LOSS ASSAY

Using a fluorescent reporter plasmid, I developed SCAMPR: a single-cell assay that infers plasmid stability in a population of host cells by measuring GFP intensity (Chapter 2). I was motivated to create this assay for several reasons. First, the throughput increase allowed for phenotyping enough progeny strains that genetic mapping should be done. Second, the sample size increase gave us better confidence in my phenotyping, minimizing sampling bias in an inherently noisy and dynamic trait. Third, single-cell resolution gave insight into loss dynamics within host populations. Having single-cell resolution was informative for constructing hypotheses about the mechanisms facilitating plasmid loss. Fourth, GFP intensity is a non-

binary readout of plasmid presence, allowing for copy number inference. Coupled with the single-cell measurements by flow cytometry, copy number heterogeneity can be assessed, which other plasmid loss assays cannot. This is powerful information especially when studying loss mechanisms of a high copy number element.

With SCAMPR validated relative to traditional plasmid loss assays, I was also able to revisit classic 2-micron fitness experiments originally performed in the 1980s. I hypothesized the plasmid is a bona fide SGE, and therefore may be driving host evolution, in part because of this established fitness cost to the host. I validated that the plasmid does indeed confer a fitness cost to the host, but also showed that SCAMPR paired with lineage tracking allows for disambiguation of plasmid fitness vs. host-plasmid incompatibility. This is a powerful tool for exploring how hosts and plasmids interact. While my study exclusively explored the mitotic stability of the 2-micron, SCAMPR could also be used with mating/sporulating populations to explore meiotic stability of the plasmid in the future.

While highly useful for this study, this same method could be expanded to studying a range of other extrachromosomal elements to clarify cost vs. stability in a range of hosts. SCAMPR would be easily adaptable for other types of plasmids or even integrating episomes in yeast, or could be brought into other microbial or tissue culture systems. In systems where GFP expression does not correlate well with DNA copy number or if the element itself does not tolerate cassette insertion, cells could instead be fixed at the final time point and probed for the element of interest via flow-FISH. Doing so would balance the benefits of working with live cells with making systems that are otherwise challenging to manipulate more accessible while maintaining throughput and large sample sizes.

Lineage tracking paired with SCAMPR is a powerful combination and can facilitate an exciting range of studies. Different fluorescent markers could be paired with polymorphic plasmids to study plasmid-plasmid competition within assorted host backgrounds. These studies would help address questions about host range, or plasmid collaboration vs. competition, and plasmid population dynamics. Done in a variety of hosts, this work could get at what determines host immunity and the breadth and specificity of that immunity. Even more excitingly, lineage tracking and SCAMPR can both be paired with experimental evolution studies. This new avenue could facilitate exciting new directions into directed host-parasite coevolution. Retrospective evolution and natural variation studies can shed light on what *has* happened during the putative yeast-plasmid arms race, and can reveal which evolutionary paths have been taken. Experimental evolution studies would open the door to understanding the range of ways this *can* play out, making this system even more powerful to understand host-parasite interactions and the repercussions of antagonistic coevolution.

5.4 NATURAL VARIATION IN ENDOGENOUS PLASMID REVEALS SOME ISOLATES HERITABLY RESTRICT 2-MICRONS

I began my investigation into natural isolates by searching for strains that are free of endogenous plasmid (Chapter 3). After testing nearly 60 strains I identified three that were plasmid-free in my hands, and were excited to see that plasmid loss is indeed a heritable trait. I next wondered what is the underlying mechanism by which plasmids are lost from these wild isolates, and what are the genetic variations that give rise to these mechanisms. I hoped to learn more about basic plasmid biology as well as how the 2-micron and host coevolve, shaping one another's genomes. To this end I characterized the loss dynamics of these wild strains using SCAMPR. I found that not only is this uncommon trait heritable, it appears to be dominant.

Additionally, it may be due to host-plasmid incompatibility, rather than variability in fitness cost between strains (see section 3.5). I infer from the plasmid loss dynamics within the host population, this is likely due to increased frequency of plasmid missegregation events.

I was surprised by, and excited to find, that the heterozygous diploid exhibits a plasmid loss phenotype. Naively, I had expected that the presence of the BY genome, which supports plasmid maintenance, would facilitate plasmid stability in the heterozygote. I initially thought that it would be more likely for the host to evolve a mutation in a permissivity factor, facilitating escape of plasmid cooption, and that this mutant would be complemented by the BY allele in the heterozygote. Instead, I did not observe complementation. This lack of complementation could be due to either 1) plasmid loss created by a dominant acting restriction factor or 2) the permissivity factor is haploinsufficient for plasmid maintenance. I suspected that the first hypothesis was more likely for the reasons detailed below; however, I cannot dismiss the possibility that the second option is true. A haploinsufficient permissivity factor seemed unlikely, however, given that permissivity factors, by definition, must also fulfill a host function. In this case the host function would be not only be haplosufficient, but completely dispensable, as Ragi strains grow without defect. I therefore moved forward expecting this locus to instead be a dominant host restriction factor. This assumption, however, can lead to downstream mapping problems if experimental design assumes dominance, when haploinsufficiency is actually the case.

A decrease in plasmid stability could be due to either fitness cost or host-plasmid incompatibility. Others had previously determined that in lab strains, rare plasmid-free cells arise in an isogenic cell population because of fitness costs. I further tested this idea with an expansion of the SCAMPR assay done in conjunction with competitive growth assays. I found that this fitness cost was the case primarily, but also that some number of plasmid loss events

are in fact due to host-plasmid incompatibility or stochastic loss. As all previous work had only looked at fitness cost in the laboratory strain, it was possible that plasmid cost is variable between yeast backgrounds. If this is the case, it could be that the notable decrease in plasmid stability in Ragi is actually due to increased fitness cost, rather than host-plasmid incompatibilities (including either permissivity or restriction factors). I suspect that the bulk of this stability difference is actually due to incompatibilities in this background, however, because the doubling times of each strain background are nearly the same, either with or without plasmid. While doubling times can not reveal very small fitness differences, the rapidity at which the plasmid is lost suggests that if loss was due to a fitness cost alone the doubling time would be detectably different. To confirm this hypothesis however, measuring fitness cost and plasmid incompatibilities via SCAMPR would remove doubt. This work is currently underway.

5.5 GENETIC MAPPING IDENTIFIES A LOCUS ASSOCIATED WITH PLASMID RESTRICTION

I wanted to map the genetic determinants of 2-micron loss in one of my wild strains (Chapter 4). I chose a QTL mapping approach by bulk segregant analysis because tetrad analysis revealed that the trait was most likely due to only a few large effect loci. Additionally, QTL mapping makes no assumptions about underlying genetic architecture for the trait of interest. This approach yielded a few loci above a LOD threshold of 4, with one significant peak of LOD ~10 on Chromosome V. As this was the most compelling locus, I focused my fine mapping efforts on this region.

I had multiple approaches available to narrow the window on my locus of interest. Ultimately I selected 3 approaches to pursue; however, in hindsight perhaps alternate approaches may

have worked better. Here I discuss the limitations of the approaches I used, possible reasons for why these methods have thus far not located variants responsible for the majority of the plasmid loss phenotype and, finally, other methods that could have be employed.

I chose three methods to pursue fine mapping: candidate gene sufficiency testing, reciprocal hemizygosity testing for necessity, and finally mitotic loss of heterozygosity as introduced by CRISPR. I discussed these approaches in depth in Chapter 4, so here I focus on why I chose them and why they may have fallen short.

5.5.1 *Sufficiency*

I tested whether Ragi candidate alleles of interest were sufficient to confer a plasmid loss phenotype to the BY lab strain haploid. I included this approach because of the possibility of most rapidly identifying the allele of interest. However, there are many reasons this approach could fail to identify the causative allele. For example, a single allele may not be sufficient to see a change in phenotype. I know that the trait appears multigenic by tetrad segregation patterns, so unless the single allele alone still gives a strong intermediate phenotype I may not confirm even a true causative allele by this method. This method makes assumptions about genetic architecture that I do not know, and also only works if the trait is truly dominant, not haploinsufficient. Finally, candidate genes were selected based on what I know about the plasmid, and the variants that are different between BY and Ragi, but shared between Ragi and Y12 (palm wine). This strategy creates a biased candidate gene list, because I suspect that not all host proteins that affect the plasmid have been identified, as well as makes assumptions about a shared origin of the heritable plasmid loss in Ragi and palm wine strains.

I had considered overcoming the biased nature of the sufficiency screen to instead look at a large list of candidates: either across the genome or a larger window of chromosome V. I considered creating CEN plasmid libraries with fragments of the Ragi genome that could be transformed into the BY strain and screened for 2-micron loss. I elected not to take this approach because of the genetic architecture concerns, namely that multiple Ragi loci may need to be present to constitute the trait, and therefore to find the gene of interest. I made the judgment that, given this risk, the amount of time required to make the vector libraries might not be worth the time investment. I also was concerned that this restriction factor could, in theory, act more broadly on expression vectors, and so the library would instead need to be integrated rather than on centromeric plasmids. Ultimately, it is possible that this strategy could have worked, or that moving more of the genetic manipulation into a lab strain background might have saved time, rather than lost it. However, I did not realize the challenges of doing genetic manipulations in Ragi at the time.

5.5.2 *Hemizygosity:*

Because some Ragi alleles may be necessary for plasmid loss, but not sufficient on their own, I reasoned that it might be easier to search for loss of my trait in knock out mutants. Some of the genes in the chromosome V locus are known to be essential (including *MMS21*, one of my candidates genes), so rather than knock out genes of interest in the haploid Ragi parent, I decided to knock out Ragi alleles in the heterozygous diploid. I reasoned that often I would generate the reciprocal BY allele knock out in the same experiment, which would provide useful controls as well. I started with the same candidate genes as in the sufficiency experiments, with the plan to begin systematically knocking out other subregions of that locus if I did not find the causative allele. However, like the sufficiency experiments this approach assumes that the trait is truly dominant, rather than the causative variant being a hypomorphic haploinsufficient allele

in Ragi. Additionally, this method is unable to detect multiple additive variants within the same region; that is to say if the locus is made up of multiple genetically linked SNPs all of which are needed for plasmid instability, this method would be unable to validate those variants. Nevertheless, I reasoned that genetic knockouts done in parallel should rapidly narrow the window of interest. I did not, however, account for the difficulties of genetic manipulations in the Ragi strain background, so this method was time, labor and resource intensive. I discovered that Ragi is not as good at homologous recombination as the laboratory strains; to obtain any transformation candidates significantly longer homology is required (1000 bp), and even then the majority of colonies that take up a selectable marker do not integrate it at the site with homology. I wonder if NHEJ may be the preferred recombination method in this strain background. It is possible the complications with genetic manipulation may in fact be linked to the plasmid-loss phenotype. The 2-micron also utilizes recombination (FLP1 mediated) as part of its copy number repair mechanism; however, even if this mechanism were suppressed in Ragi, it would still require an increased frequency of missegregation events in this background in order to explain the observed plasmid loss dynamics.

5.5.3 *Loss of heterozygosity (LOH):*

My other mapping methods were progressing more slowly than I had hoped and because I was concerned about their inherent assumptions, I chose a third strategy that would work for either a dominant or recessive haploinsufficient trait. I wanted to create a panel of heterozygous diploids that had LOH events across the chromosome V locus. Similar to the hemizygosity experiment, this approach would determine if a specific allele is necessary for the plasmid loss trait by eliminating that allele, but instead of replacing that allele with a selectable marker it would instead replace the Ragi allele with the BY allele. This replacement would address any haploinsufficiency concerns, and would work on a dominant trait as well.

I considered two methods for engineering LOH events in the heterozygote Ragi/BY strain: return to growth (RTG) and CRISPR-directed mitotic recombination. RTG creates recombination by beginning the process of meiosis, then aborting that process after recombination has taken place, allowing the cells to continue dividing mitotically still as diploids[108]. Because this is an undirected process this would facilitate recombination breakpoints throughout the entire genome. On the other hand, the CRISPR-based method creates double stranded breaks in a directed fashion, and is limited to a locus of interest. Guide RNAs are selected that target only one homolog, breaking one chromosome and allowing the cell to use the other as a repair template. Importantly, a fluorescent marker is included on one homolog end, to facilitate sorting of cells where successful mitotic recombination has taken place. I elected to take the CRISPR-driven approach because I had a specific locus on chromosome V I was most interested in fine mapping, rather than creating a large panel of randomly generated recombinants.

While the CRISPR-directed approach was created to work in a BY-RM background, the tools as available would not work in the Ragi-BY-reporter 2-micron system so this approach required significant strain construction to implement a similar conceptual method (Chapter 4). These experiments are ongoing. I am hopeful this method will best reveal the causal variant, as it is not predicated on genetic architecture requirements. Nevertheless, I am aware that this method, like those above, does require homologous recombination to succeed. However, because there is a mechanism to select for cells that have successfully eliminated one of the homologs, I am still hopeful that even if these events are less frequent in this strain background, I can still select these rare cells by flow cytometry. Because each site is chosen with specific gRNAs, large numbers of cells can be screened to find even rare events successfully and relatively quickly for sites within the locus of interest.

In addition the approaches outlined above, I had also considered other more traditional backcrossing or intercrossing methods for fine mapping. I also considered integrating selectable markers flanking my region of interest in the heterozygote, and forcing recombinants within the Chromosome V locus. At the time, however, I considered that doing the necessary strain construction, followed by rounds of sporulation, phenotyping, and sequencing library preparations would actually take longer, given that the region of interest was only ~90 kb. I anticipated that homozygous knockouts could be done in parallel and would go rapidly relative to experimental crosses. In hindsight, and given the challenges of strain construction in this background, this may not have been the case.

5.6 SPECULATIONS REGARDING THE LOCUS OF INTEREST

Perhaps I was unable to detect the allele of interest because of other incorrect assumptions. I prioritized coding change differences when pursuing candidate genes because I hypothesized antagonistic protein interaction between the host and plasmid. However, when pursuing fine mapping I included intergenic regions, in the event that a changed regulatory element was in fact responsible. However, it is possible that this approach could neglect alternative hypotheses. For example, I noted changes in the Ragi centromere relative to BY, which also falls within the chromosome V locus of interest. The centromere polymorphisms in Ragi fall within the CDEII region: one interrupts a run of A's while the other creates a run of T's. Overall the run score comparing Ragi to BY CDEII shows Ragi slightly weaker, but not a serious deficit. Although I can only speculate at this time, it is formally possible that the centromere difference itself actually underlies the plasmid loss trait. This would be an interesting future direction to explore. Perhaps a Ragi chromosome V centromere is incompatible with the 2-micron for some reason. Competition for resources would be a possibility if the Ragi centromere is weaker, but why does this not manifest as a fitness cost difference in that case? Alternatively, is the Ragi centromere somehow stronger and outcompeting the plasmid for something? But what might this be, and

how can a single centromere act dominantly to exclude a high-copy element? Regardless, non-coding elements are important to consider moving forward with fine mapping efforts.

5.7 NEW TOOLS AND NEW DIRECTIONS

Since this project first began new tools and approaches have become available for use, many of which open new and exciting doors into how yeast-2-micron coevolution can be explored. Here I discuss some resources and approaches that have come out in recent years that could change the way we study the 2-micron moving forward, and the questions I am excited to follow up on in the future.

5.8 NEW SEQUENCING RESOURCES

At the outset of this project many natural *S. cerevisiae* isolates were available to work with in the lab, many of which were also increasingly well sequenced. The SGRP collection and work from Justin Fay and Aimee Dudley and others had collected strains, and in some cases had WGS or partially sequenced (RADseq) those genomes to explore diversity and population genetics within the species. However, at that time it was not standard practice to include 2-micron reads in released genomes. A year after this study was started the 100-genomes were released, and 6 months after that the first modern genomics of the 2-micron came out as a follow up publication on those same 100 strains. These data were exciting because they followed up on work done in the ‘90s, looking at 2-micron sequence diversity, but now with the power of next generation sequencing. This study revealed a new “type” of plasmid (C type), that also clarified that the previously observed B type was in fact a recombinant of A and C types, explaining an observation in the ‘90s that one half of the plasmid might be more rapidly evolving.

I had considered looking at the discarded reads from others' previously published works, as a means to more broadly survey 2-micron diversity. This was an interesting prospect, because the combination of such a small genome and the sequences to facilitate comparative evolutionary studies could lead to hypotheses about conflict and sites of host interaction. However, this was a non-trivial undertaking, in part because the methods by which a sequencing library are prepared drastically impact plasmid presence and copy number. If others had used columns in their DNA preparations (a common practice), little to no plasmid would likely make it through to the library. Instead I opted to focus on the host genomics side of the putative conflict, and survey for a phenotype of interest in available strains.

In 2018 a much larger strain collection and associated genomes became widely available. Excitingly, this resource containing 1011 strains also included a survey of 2-micron sequences within *S. cerevisiae*. This new collection facilitates well powered comparative genomics of the 2-micron plasmid particularly well because the authors prepared all of the genomic DNA and libraries in the same way, so there is some amount of internal experimental control. This work discovered the reciprocal of recombined B type plasmids (a rare B' type), although this much less frequent type of plasmid raises questions about success and compatibility of polymorphic plasmids. Additionally a *S. paradoxus* plasmid introgression and an exceptionally divergent D type plasmid were identified. These new tools open a wide array of experimental and bioinformatics questions that I am excited to pursue in the future. What determines plasmid host range or success? The SCAMPR assay paired with differently tagged polymorphic plasmids could begin to address plasmid fitness relative to one another, and in the contexts of different host backgrounds.

This same publication also found that Y9 and Y12 are plasmid negative (independently validating my observations), but also that there are other monophyletic branches of the tree that do not contain endogenous plasmid by WGS (such as the French Dairy clade). I hypothesize that these lineages that are all missing the 2-micron are more likely to have a heritable basis for plasmid loss, while single plasmid-free tips might be due to stochastic loss, strain handling or sequencing library artifacts. Given the number of mating opportunities these strains have experienced since their last common ancestor, I would be surprised if none of the related strains would have had plasmid reintroduced in that time. I favor the hypothesis that the common ancestor instead gained a heritable means to restrict the plasmid, leading to these plasmid-free sub groups. To determine if strains are arriving at similar or different means to lose the 2-micron, it will be exciting to look more broadly within *S. cerevisiae* at the restriction mechanisms that have naturally evolved and at what (if any) fitness benefit and cost tradeoffs exist for these molecular mechanisms.

Utilizing these groups and their plasmid-containing outgroups could lead to better powered comparative genomics to hone in on causative loci. If comparative genomics are not enough to identify interesting loci to follow up, this is still a powerful tool to use in conjunction with mapping options in genetic crosses. Fixed SNPs within a locus of interest that are not shared with plasmid-containing outgroups would be most suspect. Additionally, knowing which closely related strains do and do not have plasmid could allow for better resolution QTN mapping[109]; I could perform crosses between much more closely related strains that have different plasmid stability phenotypes, the nucleotide-level variant that is responsible may be easier to determine.

It would be interesting to see if regions of the host genome share similar phylogenomics as the plasmids themselves. Because all of these variants are *within* species, there is gene flow amongst the strains; for this reason the strain tree might not reflect the same tree as a region of

the genome (a portion of chromosome V for instance). Are there host alleles that correlate with which type of plasmid is endogenously found in a host? It would be interesting to explore if there are host genes that match the tree of plasmid presence and genetic variation, even if the rest of the genome does not share that pattern.

Beyond *S. cerevisiae*, the Y1000+ project is an ongoing and collaborative effort to explore yeast species diversity. This work, in addition to other resources curated by the Wolfe and Hittinger labs among others, provide exciting future opportunities for exploring both host genome and plasmid diversity across a much longer evolutionary timeline. The Y1000+ project in particular does not use columns in their sequencing preparations, and is a large collection being handled consistently across species, so it will be exciting to mine the 2-micron data once available. Not only will these sequences provide the opportunity to compare plasmid evolution down different lineages, wholly plasmid-free species might provide exciting opportunities to explore how hosts can “win” against an SGE. Plasmid horizontal transfer events or introgression events could also reveal key insights into host range and permissivity factor flexibility and constraints. Finally, as these other species gain interest for biotechnology purposes (due to novel metabolic pathways and products), these diverged endogenous plasmids could be harnessed to make expression vectors, just as the *S. cerevisiae* 2-micron was in the 80’s. These tools, in conjunction with CRISPR technology, could facilitate emergent model systems for production of rare bioproducts and other studies.

5.9 EVOLUTIONARY ANALYSIS

Genetic conflicts can be explored through two complementary angles: one, to identify an antagonistic molecular interaction, then search for signs of coevolutionary impact (such as evidence of positive selection). Or, two: Search for evidence of positive selection, then search for the antagonist driving that rapid evolution. In my study I began with a known antagonist, and

hoped to identify the interface at which conflict is playing out (e.g., a restriction factor). While other studies have taken the second approach, systematically looking for genes with signatures of rapid evolution, the yeast genome is still large enough that tying these results to a specific antagonist is challenging without complete knowledge of the plasmid lifecycle and host immunity mechanisms.

However, the 2-micron plasmid is a uniquely tiny genome. A greater diversity of plasmid sequences opens up the possibility of evolutionary studies from the plasmid's side of the arms race. It will be exciting to perform evolutionary analyses of the plasmid, then identify which host components may be driving this evolution. With a specific region of plasmid as bait, it will be easier to interrogate which host components interact, and look for drivers of any detected positive selection.

5.10 PLASMIDS AS A SOURCE OF HOST BIOLOGICAL NOVELTY

Interestingly, I and others have observed that in some species (although not *S. cerevisiae*) 2-micron genes have been integrated into the host genome. Although not yet tested functionally, these integration events seem to correlate with species that are not known to harbor a full endogenous plasmid. It would be interesting to explore if these genes, once integrated into the host, are being coopted and utilized as plasmid restriction factors. Perhaps this correlation is in part due to sampling bias for some of these species (some strains have only a single type strain sequenced), or maybe this correlation hints at one route that hosts may "win" against their parasites. Alternatively, sometimes genetic conflict can facilitate biological novelty on other fronts; perhaps these coopted genes are instead being neofunctionalized for completely new roles in the host. Functional and evolutionary studies of these integrated genes will be exciting, especially as compared to their plasmid-based counterparts in neighboring species. An exciting

possibility would be re-creating these gene capture events then experimentally evolving yeast in laboratory to see if and how the host can use integrated plasmid genes.

5.11 OTHER METHODS FOR IDENTIFYING PLASMID PERMISSIVITY AND RESTRICTION FACTORS

Here, for reasons detailed previously, I elected to use a natural variation approach to understand host-plasmid interactions. However, there are other methods available, as others have explored previously. Where I focused on genetics-based approaches, protein and RNA based approaches could also reveal interesting host-plasmid biology. Transcriptional profiling via RNAseq between plasmid-positive or plasmid-free cells could reveal more about the host response to the 2-micron. Combining this with natural isolates could reveal if some wild strains regulate these responses differently, or if new or different genes are up regulated in strains that successfully limit the plasmid.

Protein-protein interactions have been explored by others, with a number of papers having come from the Jayaram lab and others detailing host proteins that have evidence of plasmid interactions either via DNA or protein-protein interactions. These studies appear to have been largely focused on single proteins. High throughput proteomics between different strain backgrounds could reveal either permissivity factors that fail to associate in wild isolates, or perhaps novel interactors in these natural strains would identify restriction factors. While individual proteins that have undergone rigorous validation for plasmid interactions are invaluable for understanding the plasmid lifecycle, it is possible that a higher throughput screen of all observed interactions could also help paint a more comprehensive picture, and if available publicly would be a convenient resource for helping select candidate genes from genetic mapping experiments. However, because of the high copy number of the plasmid, false-positive artifacts are a very real concern by -omics approaches. Crosslinking in experiments maximizes

the chances of pulling down nearby, but functionally unrelated, proteins at substoichiometric levels. So while these sorts of data would be helpful public resources, additional experiments would certainly be required to conclusively determine plasmid-related functions.

Experimental evolution could provide an exciting new angle for exploring host-plasmid interactions as well as coevolutionary repercussions thereof, given the fitness cost of the plasmid. Most experimental evolution previously performed in budding yeast has in fact been done in laboratory strain backgrounds known to contain the plasmid. Perhaps these datasets could be retrospectively mined to look for plasmid changes. Are there specific conditions in which is it more or less costly to the host, or where plasmid-free isolates sweep a population? While 1-3% is moderate compared to some fitness gains that somatic mutations achieve in these strongly selective experiments, understanding plasmid dynamics in these experiments broadly might inform how this fitness cost occurs. Presumably this cost is due to replication stress, based on plasmid overabundance experiments; however, in a typical plasmid lifecycle perhaps this is not the case. Experimental evolution could reveal when the plasmid is more costly and likely to be lost, or perhaps if there is ever a cryptic benefit to the plasmid, one might see a sweep of plasmid positive cells. The Murray lab and others have reported that in some mutant strain backgrounds replicating more slowly can benefit the host, presumably buying the cell time to correct other errors that would otherwise be deadly in the already sick background[110]. In this case having the 2-micron might actually provide some host benefit. More specific environmental conditions could reveal other cryptic benefits if the plasmid is ever selected for. One caveat of retrospective studies of 2-micron, however, is its sensitivity to sequencing preparation methods. One would need to work closely with the original authors who generated that data to ensure that DNA and library prep methods do not bias the observed results.

Another strategy to understand how evolution of permissivity or restriction factors can impede 2-micron success is to use experimental evolution to evolve ‘escape’ from 2-micron plasmids. Indeed, such an experimental evolution strategy is a robust way to acquire mutations that provide protection against pathogenic viruses, for example. However, the fitness consequences of 2-micron plasmids may be too subtle for such studies. Exploring current datasets for more information about 2-micron stability may address this concern. However, even if the fitness burden of the endogenous plasmid is too subtle on its own, the SCAMPR reporter plasmid could be modified to include a titratable fitness cost to the host. For example, if a tunable promoter was driving expression of a high-cost protein, one could experimentally “turn up” fitness cost of having the 2-micron during evolution experiments. By increasing plasmid burden to the host in this way, perhaps evolution of immunity factors could be driven in the laboratory. One could chart the successful mechanisms employed by the host to overcome a mounting plasmid burden.

Genetic conflict can drive biological innovation, not only through directly providing selective antagonistic pressure, but sometimes through being an exogenous source of variation. Do different strains harboring 2-micron plasmids have different evolutionary routes available when presented with the same environmental stressors? Experimental evolution often aims to discover how much of evolution is stochastic and how much is deterministic and repeatable given the same selective pressures. Plasmid negative and positive strains could be evolved in the same environmental conditions to explore how much conflict might be influencing this determinism. These experiments would be even more compelling if a screen for conditionally essential genes were to reveal which host functions the 2-micron taxes or benefits, and could be used to further explore conflict in the context of gene-by-environment interactions.

5.12 Genetic screens to explore cryptic benefits, permissivity actors and weak restrictors of the 2-micron plasmid

Although I speculate that the plasmid is unlikely to ever benefit the host, given the genes present on the plasmid itself, and its self-recovery mechanisms, I cannot entirely exclude the possibility that the plasmid actually is sometimes beneficial to the host. It is surprising that an element without an addiction or integration method would be so infrequently lost in a host cell, particularly given how ancient I believe the 2-micron/yeast coevolution has been going on. One approach that could reveal roles of the 2-micron plasmid within the cell would be to explore conditionally essential genes within the host. Are there any host genes that are non-essential in the presence of the plasmid that become essential in a plasmid-negative background? This experiment could reveal redundancies or context specific benefits of the plasmid. On the other hand, this same experiment could reveal genes that are non-essential in a plasmid-free cell, but are required in the presence of the plasmid. This experiment could reveal weak restrictors of the 2-micron plasmid, host genes that keep the plasmid burden at a manageable level, and could reveal more about where the host and plasmid interact at the molecular level and what molecular paths are taxed by the 2-micron.

Excitingly, in 2017 the Kornmann group published a new method for doing saturating transposon mutagenesis in a single step in *S. cerevisiae* - SATAY. While not yet easily used in some wild strain backgrounds, a rotation student Sam Hart worked with me during his rotation to get SATAY working in 2-micron positive and negative laboratory strains. His initial experimental results did not reach saturation, unfortunately, making his data set too noisy to interpret conclusively; however, I believe this will be a fruitful method moving forward with concrete troubleshooting steps needed to get the experimental numbers required for saturation.

Furthermore it will be interesting ultimately to explore if and how the list of conditionally essential genes compare between differing strain backgrounds, such as natural isolates.

S. cerevisiae in particular is a powerful species for doing genetic screens, largely due to the number of tools and resources available from the established research community. One of those resources is the yeast deletion collection, where most non-essential ORFs have been knocked out of the BY laboratory background strain. Because this strain background contains an endogenous 2-micron plasmid, most of the deletion collection should as well. I hypothesized that if any of these non-essential genes were plasmid permissivity factors, that perhaps the knockout line would have lost the endogenous plasmid. As such I set out to screen the deletion collection with the help of Cynthis Wong, an undergraduate researcher, looking for any deletion collection lines that had lost the 2-micron plasmid. Because of how the deletion collection was created, and the number of subsequent passages since creation, I also knew stochastic bottleneck events would be a large confounder of this screen. So I planned to use SCAMPR to validate candidate permissivity factors. While Cynthis screened the bulk of the deletion collection by PCR and identified several candidate lines without 2-micron, she was unable to perform candidate validation prior to leaving for medical school. This screen was limited in that it can only identify non-essential host factors that are permissivity factors for plasmid maintenance (and many are known to be host-essential as well), but is an excellent example of the resources available in the yeast community that are often unavailable in other host-SGE model systems. Interestingly, Cynthis found a surprising number of nuclear-encoded mitochondrial gene knockouts were plasmid-negative in her screen. While I have not yet followed up on this further, it is interesting as Ragi is known to have a mitochondrial genome (mitotype) that is fitness-beneficial under some stressful conditions. Exploring how mitochondria and plasmids might indirectly interact within a host cell affecting fitness is an interesting direction, particularly given the non-Mendelian inheritance patterns of both.

So far all of my studies have focused on mitotic cost and stability of the 2-micron plasmid. However, it would be interesting to explore similar questions through meiosis. The 2-micron undergoes non-Mendelian inheritance, with all progeny in a cross typically inheriting the plasmid even when only one parent was infected. SCAMPR could be paired with a fluorescent marker of sporulation: successfully sporulated cells could be measured by flow cytometry and their plasmid status could be assessed. Perhaps the 2-micron is beneficial to the host during rounds of sex and meiosis, while confers a cost during clonal mitotic growth. It will be interesting to leverage SCAMPR to study meiotic costs and benefits of the 2-micron on their host cells.

5.13 EXPERIMENTALLY DETERMINING AVAILABLE PATHS OF HOST EVOLUTION

Many of the host-encoded proteins known to interact with the plasmid, such as cohesin and the replication origin licensing machinery, are essential for host survival as well. However, some of these permissivity factor proteins may be able to sample sequence space that may allow them to escape plasmid hijacking, while others may be tightly constrained or trapped at a local maxima for host functionality. It would be interesting to know in the case of plasmid-interacting factors that do not show signatures of rapid evolution, what their fitness landscape looks like. In order to understand how plasmid and host shape one another's evolution it is important to understand how these genomes interact within the cell, as well as the constraints on both for self-function and environmental interactions. While some separation of function variants may have naturally arisen in the evolution of plasmid-free isolates, exploring what has already happened only gives a small window into what possibilities might happen and which are viable. To systematically explore the fitness landscape for permissivity factors, and potential plasmid-evading mutations, one could leverage deep mutational scanning (DMS) on known plasmid permissivity factors. All possible amino acid substitutions of a permissivity factor could be tested

in both plasmid-free and plasmid-containing strains, and corresponding plasmid stability could be measured by SCAMPR in the latter.

5.14 IN SUMMARY

While biochemistry and traditional genetics have gleaned the information we have about the 2-micron's typical lifecycle, relatively few studies have explored the evolution of the plasmid or the host's evolution in the context of plasmid antagonism[33,96,97,111]. Exploring this putative coevolution could yield information both generally about how hosts coevolve with chronic genetic parasites, but also may reveal details about the mechanisms underlying how host and parasite interact at the molecular and functional levels. Budding yeast and 2-microns provide a compelling, and increasingly powerful, model system for understanding how genomes in conflict evolve. New resources increase the breadth of evolutionary questions that can be asked on both short and long time scales; as this has played out potentially differently in a multitude of lineages across a wide clade of budding yeast species potentially with different outcomes. New tools and molecular methods facilitate deep functional and mechanistic studies in *S. cerevisiae* and beyond. In this work I used a natural, often overlooked, selfish element in one of the best-studied eukaryotes in the world to build a promising new model system for understanding host-parasite coevolution.

BIBLIOGRAPHY

1. Burt, A. and Trivers, R. *Genes in Conflict*. Harvard University Press. Harvard Univeristy Press (2006).
2. Dawkins, R. *The selfish gene*. Oxford University Press (1989).
3. McLaughlin, R. N. & Malik, H. S. Genetic conflicts: The usual suspects and beyond. *Journal of Experimental Biology* (2017). doi:10.1242/jeb.148148
4. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annu. Rev. Biochem.* (2013). doi:10.1146/annurev-biochem-072911-172315
5. Colomer-Lluch, M., Ruiz, A., Moris, A. & Prado, J. G. Restriction Factors: From Intrinsic Viral Restriction to Shaping Cellular Immunity Against HIV-1. *Frontiers in immunology* (2018). doi:10.3389/fimmu.2018.02876
6. Rowley, P. A., Ho, B., Bushong, S., Johnson, A. & Sawyer, S. L. XRN1 Is a Species-Specific Virus Restriction Factor in Yeasts. *PLOS Pathog.* **12**, e1005890 (2016).
7. Rowley, P. A. The frenemies within: viruses, retrotransposons and plasmids that naturally infect Saccharomyces yeasts. *Yeast* (2017). doi:10.1002/yea.3234
8. Klein, S. J. & O'Neill, R. J. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Research* (2018). doi:10.1007/s10577-017-9569-5
9. Molaro, A. & Malik, H. S. Hide and seek: How chromatin-based pathways silence retroelements in the mammalian germline. *Current Opinion in Genetics and Development* (2016). doi:10.1016/j.gde.2015.12.001
10. Unterholzner, S. J., Poppenberger, B. & Rozhon, W. Toxin-antitoxin systems: Biology, identification, and application. *Mob. Genet. Elements* (2013). doi:10.4161/mge.26219
11. Bravo Núñez, M. A., Nuckolls, N. L. & Zanders, S. E. Genetic Villains: Killer Meiotic Drivers. *Trends in Genetics* (2018). doi:10.1016/j.tig.2018.02.003
12. Larracuente, A. M. & Presgraves, D. C. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics* (2012). doi:10.1534/genetics.112.141390
13. Zanders, S. E. *et al.* Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *Elife* (2014). doi:10.7554/eLife.02630.001
14. Chmátal, L., Schultz, R. M., Black, B. E. & Lampson, M. A. Cell Biology of Cheating-Transmission of Centromeres and Other Selfish Elements Through Asymmetric Meiosis. *Progress in molecular and subcellular biology* (2017). doi:10.1007/978-3-319-58592-5_16
15. Chmátal, L. *et al.* Centromere strength provides the cell biological basis for meiotic drive

and karyotype evolution in mice. *Curr. Biol.* (2014). doi:10.1016/j.cub.2014.08.017

16. Daugherty, M. D. & Malik, H. S. Rules of Engagement: Molecular Insights from Host-Virus Arms Races. *Annu. Rev. Genet.* (2012). doi:10.1146/annurev-genet-110711-155522
17. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution [33]. *Nature* (1977). doi:10.1038/267275a0
18. Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503 (2000).
19. Hsu, T. H. & Spindler, K. R. Identifying host factors that regulate viral infection. *PLoS Pathog.* (2012). doi:10.1371/journal.ppat.1002772
20. Werren, J. H. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci. U. S. A.* (2011). doi:10.1073/pnas.1102343108
21. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–9 (2000).
22. Elde, N. C., Child, S. J., Geballe, A. P. & Malik, H. S. Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature* (2009). doi:10.1038/nature07529
23. Mitchell, P. S. *et al.* Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe* (2012). doi:10.1016/j.chom.2012.09.005
24. Henikoff, S. & Malik, H. S. Centromeres: Selfish drivers. *Nature* (2002). doi:10.1038/417227a
25. Malik, H. S. & Henikoff, S. Major Evolutionary Transitions in Centromere Complexity. *Cell* (2009). doi:10.1016/j.cell.2009.08.036
26. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–41 (2009).
27. Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A. & Bai, F.-Y. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol. Ecol.* **21**, 5404–5417 (2012).
28. Scannell, D. R., Butler, G. & Wolfe, K. H. Yeast genome evolution - The origin of the species. *Yeast* (2007). doi:10.1002/yea.1515
29. Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15**, 1456–61 (2005).
30. Hyma, K. E. & Fay, J. C. Mixing of vineyard and oak-tree ecotypes of Saccharomyces cerevisiae in North American vineyards. *Mol. Ecol.* **22**, 2917–30 (2013).

31. Fay, J. C. Tapping into yeast diversity. *Molecular Ecology* (2012). doi:10.1111/mec.12035
32. Cromie, G. A. *et al.* Genomic sequence diversity and population structure of Saccharomyces cerevisiae assessed by RAD-seq. *G3 (Bethesda).* **3**, 2163–71 (2013).
33. Peter, J. *et al.* Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* (2018). doi:10.1038/s41586-018-0030-5
34. Strope, P. K. *et al.* The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* (2015). doi:10.1101/gr.185538.114
35. The awesome power of yeast genetics - practical approaches and recipes for success: Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2000 Edition. *J. Cell Sci.* (2001).
36. Rizvi, S. M. A., Prajapati, H. K. & Ghosh, S. K. The 2 micron plasmid: a selfish genetic element with an optimized survival strategy within Saccharomyces cerevisiae. *Current Genetics* (2018). doi:10.1007/s00294-017-0719-2
37. Dobson, M. J. *et al.* The 2 micron plasmid causes cell death in Saccharomyces cerevisiae with a mutation in Ulp1 protease. *Mol. Cell. Biol.* **25**, 4299–310 (2005).
38. Mehta, S. *et al.* The 2 micron plasmid purloins the yeast cohesin complex: A mechanism for coupling plasmid partitioning and chromosome segregation? *J. Cell Biol.* **158**, 625–637 (2002).
39. Dobson, M. J., Yull, F. E., Molina, M., Kingsman, S. M. & Kingsman, A. J. Reconstruction of the yeast 2 μm plasmid partitioning mechanism. *Nucleic Acids Res.* (1988). doi:10.1093/nar/16.14.7103
40. Pinder, J. B., McQuaid, M. E. & Dobson, M. J. Deficient sumoylation of yeast 2-micron plasmid proteins Rep1 and Rep2 associated with their loss from the plasmid-partitioning locus and impaired plasmid inheritance. *PLoS One* **8**, e60384 (2013).
41. Huang, C.-C., Chang, K.-M., Cui, H. & Jayaram, M. Histone H3-variant Cse4-induced positive DNA supercoiling in the yeast plasmid has implications for a plasmid origin of a chromosome centromere. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13671–6 (2011).
42. Velmurugan, S., Ahn, Y. T., Yang, X. M., Wu, X. L. & Jayaram, M. The 2 micrometer plasmid stability system: analyses of the interactions among plasmid- and host-encoded components. *Mol. Cell. Biol.* **18**, 7466–77 (1998).
43. Xiao, W., Pelcher, L. E. & Rank, G. H. Evidence for cis- and trans-acting element coevolution of the 2-microns circle genome in Saccharomyces cerevisiae. *J. Mol. Evol.* **32**, 145–52 (1991).

44. Xie, Y., Pelcher, L. E. & Rank, G. H. Chimeric evolution of the 2-microns genome in Saccharomyces cerevisiae. *J. Mol. Evol.* **38**, 363–8 (1994).
45. Hittinger, C. T. *et al.* Genomics and the making of yeast biodiversity. *Current Opinion in Genetics and Development* (2015). doi:10.1016/j.gde.2015.10.008
46. Utatsu, I., Sakamoto, S., Imura, T. & Toh-e, A. Yeast plasmids resembling 2 micron DNA: regional similarities and diversities at the molecular level. *J. Bacteriol.* **169**, 5537–45 (1987).
47. Blaisonneau, J., Sor, F., Cheret, G., Yarrow, D. & Fukuhara, H. A circular plasmid from the yeast Torulaspora delbrueckii. *Plasmid* **38**, 202–9 (1997).
48. Chen, X. J., Cong, Y. S., Wesolowski-Louvel, M., Li, Y. Y. & Fukuhara, H. Characterization of a circular plasmid from the yeast Kluyveromyces waltii. *J. Gen. Microbiol.* **138**, 337–345 (1992).
49. Murray, J. A. H., Cesareni, G. & Argos, P. Unexpected divergence and molecular coevolution in yeast plasmids. *J. Mol. Biol.* **200**, 601–607 (1988).
50. Broach, J. R., Guarascio, V. R. & Jayaram, M. Recombination within the yeast plasmid 2μ circle is site-specific. *Cell* **29**, 227–234 (1982).
51. Broach, J. R. The yeast plasmid 2μ circle. *Cell* **28**, 203–204 (1982).
52. BROACH, J. R., LI, Y.-Y., WU, L.-C. C. & JAYARAM, M. Vectors for High-Level, Inducible Expression of Cloned Genes in Yeast. in *Experimental Manipulation of Gene Expression* (1983). doi:10.1016/b978-0-12-372380-2.50009-3
53. Velmurugan, S., Ahn, Y.-T., Yang, X.-M., Wu, X.-L. & Jayaram, M. The 2μm Plasmid Stability System: Analyses of the Interactions among Plasmid- and Host-Encoded Components. *Mol. Cell. Biol.* **18**, 7466–7477 (1998).
54. Huang, L. C., Wood, E. A. & Cox, M. M. A bacterial model system for chromosomal targeting. *Nucleic Acids Research* (1991). doi:10.1093/nar/19.8.1978
55. Chan, K.-M., Liu, Y.-T., Ma, C.-H., Jayaram, M. & Sau, S. The 2 micron plasmid of Saccharomyces cerevisiae: a miniaturized selfish genome with optimized functional competence. *Plasmid* **70**, 2–17 (2013).
56. Zakian, V. A., Brewer, B. J. & Fangman, W. L. Replication of Each Copy of the Yeast 2 Micron DNA Plasmid Occurs during the S Phase. *Cell* **17**, 923–934 (1979).
57. Huang, C.-C. Functional interactions of chromosome segregation factors with the 2 micron plasmid : possible evolutionary link between the plasmid portioning locus and the budding yeast centromere. (2011).
58. Mead, D. J., Gardner, D. C. & Oliver, S. G. The yeast 2 micron plasmid: strategies for the

survival of a selfish DNA. *Mol. Gen. Genet.* **205**, 417–21 (1986).

59. Futcher, A. B. & Cox, B. S. Maintenance of the 2 microns circle plasmid in populations of Saccharomyces cerevisiae. *J. Bacteriol.* **154**, 612–22 (1983).
60. Huberman, J. A., Spotila, L. D., Nawotka, K. A., El-Assouli, S. M. & Davis, L. R. The in vivo replication origin of the yeast 2μm plasmid. *Cell* **51**, 473–481 (1987).
61. Jayaram, M., Li, Y.-Y. & Broach, J. R. The yeast plasmid 2μ circle encodes components required for its high copy propagation. *Cell* **34**, 95–104 (1983).
62. Jayaram, M., Li, Y. Y. & Broach, J. R. The yeast plasmid 2μ circle encodes components required for its high copy propagation. *Cell* (1983). doi:10.1016/0092-8674(83)90139-3
63. Velmurugan, S. Partitioning of the 2-microm Circle Plasmid of Saccharomyces cerevisiae: Functional Coordination with Chromosome Segregation and Plasmid-encoded Rep Protein Distribution. *J. Cell Biol.* **149**, 553–566 (2000).
64. Wong, M. C. V. L., Scott-Drew, S. R. S., Hayes, M. J., Howard, P. J. & Murray, J. A. H. RSC2, encoding a component of the RSC nucleosome remodeling complex, is essential for 2 microm plasmid maintenance in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **22**, 4218–29 (2002).
65. Prajapati, H. K., Rizvi, S. M. A., Rathore, I. & Ghosh, S. K. Microtubule-associated proteins, Bik1 and Bim1, are required for faithful partitioning of the endogenous 2 micron plasmids in budding yeast. *Mol. Microbiol.* (2017). doi:10.1111/mmi.13608
66. Ouspenski, I. I., Elledge, S. J. & Brinkley, B. R. New yeast genes important for chromosome integrity and segregation identified by dosage effects on genome stability. *Nucleic Acids Res.* **27**, 3001–8 (1999).
67. Futcher, A. B. & Cox, B. S. Copy number and the stability of 2-micron circle-based artificial plasmids of Saccharomyces cerevisiae. *J. Bacteriol.* **157**, 283–90 (1984).
68. Velmurugan, S., Yang, X., Jayaram, M., Mehta, S. & Voziyanov, Y. The 2μm Plasmid of Saccharomyces cerevisiae. in *Plasmid Biology* (2014). doi:10.1128/9781555817732.ch14
69. Volkert, F. C. & Broach, J. R. Site-specific recombination promotes plasmid amplification in yeast. *Cell* (1986). doi:10.1016/0092-8674(86)90879-2
70. Som, T., Armstrong, K. A., Volkert, F. C. & Broach, J. R. Autoregulation of 2μm circle gene expression provides a model for maintenance of stable plasmid copy levels. *Cell* (1988). doi:10.1016/0092-8674(88)90528-4
71. Chen, X. L., Reindle, A. & Johnson, E. S. Misregulation of 2 m Circle Copy Number in a SUMO Pathway Mutant. *Mol. Cell. Biol.* **25**, 4311–4320 (2005).
72. Ma, C. H. *et al.* A Flp-SUMO hybrid recombinase reveals multi-layered copy number

control of a selfish DNA element through post-translational modification. *PLoS Genet.* (2019). doi:10.1371/journal.pgen.1008193

73. Zhao, X., Wu, C. Y. & Blobel, G. Mlp-dependent anchorage and stabilization of a desumoylating enzyme is required to prevent clonal lethality. *J. Cell Biol.* (2004). doi:10.1083/jcb.200405168
74. Bennett, P. M. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. in *British Journal of Pharmacology* (2008). doi:10.1038/sj.bjp.0707607
75. Schmitt, M. J. & Reiter, J. Viral induced yeast apoptosis. *Biochim. Biophys. Acta* **1783**, 1413–7 (2008).
76. Maine, G. T., Sinha, P. & Tye, B. K. Mutants of S. cerevisiae defective in the maintenance of minichromosomes. *Genetics* **106**, 365–85 (1984).
77. Ludwig, D. L. & Bruschi, C. V. The 2-μm plasmid as a nonselectable, stable, high copy number yeast vector. *Plasmid* **25**, 81–95 (1991).
78. Lee, C., Kim, J., Shin, S. G. & Hwang, S. Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli. *J. Biotechnol.* (2006). doi:10.1016/j.jbiotec.2005.11.014
79. Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* (2008). doi:10.1038/nmeth.1234
80. Tsalik, E. L. & Gartenberg, M. R. Curing Saccharomyces cerevisiae of the 2 micron plasmid by targeted DNA damage. *Yeast* **14**, 847–52 (1998).
81. Ma, C.-H. *et al.* Temporal sequence and cell cycle cues in the assembly of host factors at the yeast 2 micron plasmid partitioning locus. *Nucleic Acids Res.* **41**, 2340–53 (2013).
82. Nishimura, A. *et al.* Mitochondrial cysteinyl-tRNA synthetase is expressed via alternative transcriptional initiation regulated by energy metabolism in yeast cells. *J. Biol. Chem.* (2019). doi:10.1074/jbc.RA119.009203
83. Suzuki, Y. *et al.* The green monster process for the generation of yeast strains carrying multiple gene deletions. *J. Vis. Exp.* e4072 (2012). doi:10.3791/4072
84. Labunskyy, V. M. *et al.* The insertion Green Monster (iGM) method for expression of multiple exogenous genes in yeast. *G3 Genes, Genomes, Genet.* (2014). doi:10.1534/g3.114.010868
85. Lauer, S. *et al.* Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. *PLoS Biol.* (2018). doi:10.1371/journal.pbio.3000069
86. Zhu, J. *et al.* Single-cell based quantitative assay of chromosome transmission fidelity.

G3 Genes, Genomes, Genet. (2015). doi:10.1534/g3.115.017913

87. Kao, K. C. & Sherlock, G. Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. *Nat. Genet.* (2008). doi:10.1038/ng.280
88. Hailey, D. W., Davis, T. N. & Muller, E. G. D. Fluorescence resonance energy transfer using color variants of green fluorescent protein. *Methods Enzymol.* (2002). doi:10.1016/S0076-6879(02)51840-1
89. Amberg, D. C., Burke, D. J. & Strathern, J. N. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition*. *A Cold Spring Harbor Laboratory Course Manual* (2005).
90. Rowley, P. A., Ho, B., Bushong, S., Johnson, A. & Sawyer, S. L. XRN1 Is a Species-Specific Virus Restriction Factor in Yeasts. *PLoS Pathog.* (2016). doi:10.1371/journal.ppat.1005890
91. Krastanova, O., Hadzhitodorov, M. & Pesheva, M. Ty elements of the yeast saccharomyces cerevisiae. *Biotechnol. Biotechnol. Equip.* (2005). doi:10.1080/13102818.2005.10817272
92. Maxwell, P. H. & Curcio, M. J. Host factors that control long terminal repeat retrotransposons in Saccharomyces cerevisiae: Implications for regulation of mammalian retroviruses. *Eukaryotic Cell* (2007). doi:10.1128/EC.00092-07
93. Toh-e, A., Utatsu, I., Utsunomiya, A., Sakamoto, S. & Imura, T. *Extrachromosomal Elements in Lower Eukaryotes*. (Springer US, 1987). doi:10.1007/978-1-4684-5251-8
94. Wickner, R. B. Killer yeasts. *Current topics in medical mycology* (1985).
95. Wickner, R. Double-Stranded and Single-Stranded RNA Viruses of Saccharomyces Cerevisiae. *Annu. Rev. Microbiol.* (1992). doi:10.1146/annurev.micro.46.1.347
96. Strope, P. K. *et al.* 2μ plasmid in Saccharomyces species and in Saccharomyces cerevisiae. *FEMS Yeast Res.* **15**, (2015).
97. Xiao, W., Pelcher, L. E. & Rank, G. H. Sequence diversity of yeast 2 μm RAF gene and its co-evolution with STB and REP1. *Gene* **101**, 75–80 (1991).
98. Hose, J. *et al.* Dosage compensation can buffer copy-number variation in wild yeast. *Elife* **4**, (2015).
99. Pavelka, N. *et al.* Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. doi:10.1038/nature09529
100. Edwards, M. D. & Gifford, D. K. High-resolution genetic mapping with pooled sequencing. *BMC Bioinforma. 2012 136* **13**, S8 (2012).

101. Sawyer, S. L. & Malik, H. S. Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17614–9 (2006).
102. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* (2003). doi:10.1038/nature02026
103. Zhao, X. & Blobel, G. A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proc. Natl. Acad. Sci. U. S. A.* (2005). doi:10.1073/pnas.0500537102
104. Bermúdez-López, M. *et al.* The Smc5/6 complex is required for dissolution of DNA-mediated sister chromatid linkages. *Nucleic Acids Res.* (2010). doi:10.1093/nar/gkq546
105. Wilkening, S. *et al.* An evaluation of high-throughput approaches to QTL mapping in Saccharomyces cerevisiae. *Genetics* (2014). doi:10.1534/genetics.113.160291
106. Romano, G. H. *et al.* Different sets of QTLs influence fitness variation in yeast. *Mol. Syst. Biol.* (2010). doi:10.1038/msb.2010.1
107. Sadhu, M. J., Bloom, J. S., Day, L. & Kruglyak, L. CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science* **352**, 1113–6 (2016).
108. Laureau, R. *et al.* Extensive Recombination of a Yeast Diploid Hybrid through Meiotic Reversion. *PLoS Genet.* (2016). doi:10.1371/journal.pgen.1005781
109. She, R. & Jarosz, D. F. Mapping Causal Variants with Single-Nucleotide Resolution Reveals Biochemical Drivers of Phenotypic Change. *Cell* (2018). doi:10.1016/j.cell.2017.12.015
110. Fumasoni, M. & Murray, A. W. The evolutionary plasticity of chromosome metabolism allows adaptation to DNA replication stress. *bioRxiv* (2019). doi:10.1101/770859
111. Xiao, W., Pelcher, L. E. & Rank, G. H. Evidence forCis- andTrans-acting element coevolution of the 2-μm circle genome inSaccharomyces cerevisiae. *J. Mol. Evol.* **32**, 145–152 (1991).

VITA

Originally from California, Michelle Hays spent most of her formative years in Colorado. She is the first in her family to pursue a college degree, the first to attend graduate school and the first to complete a PhD. She obtained her B.S. in Microbiology and German Language, Literature and Culture from Colorado State University in 2005. Prior to beginning graduate school she worked both in biotech and in academic research labs, most notably at the Institute for Systems Biology from 2009-2013 in the labs of Dr. Tim Galitski and Dr. Aimee Dudley. From her time at ISB she is author on nine publications. It was during this time she also began volunteering at North Seattle Community College helping facilitate organic chemistry laboratory courses and guest lecturing. She fell in love with mentoring and teaching, just as she fell in love with research. It was this passion that inspired her to earn her PhD in the lab of Dr. Harmit Malik, in hopes of eventually obtaining a faculty position where she can mentor others in both the laboratory and classroom. To that end, Dr. Malik has been supportive of her ongoing outreach and teaching endeavors in graduate school, including the creation of a new teaching and pedagogy training program for University of Washington graduate students. This program has subsequently gained departmental support and NSF funding and will continue to run after she leaves UW. She will be starting an academic postdoctoral fellowship position at Stanford in the beginning of 2020, in the laboratory of Dr. Gavin Sherlock.