

Winter 12-15-2018

# Genetic Basis of Thermal Divergence in *Saccharomyces* species

Xueying C. Li

*Washington University in St. Louis*

Follow this and additional works at: [https://openscholarship.wustl.edu/art\\_sci\\_etds](https://openscholarship.wustl.edu/art_sci_etds)

 Part of the [Bioinformatics Commons](#), [Developmental Biology Commons](#), [Evolution Commons](#),  
and the [Genetics Commons](#)

---

## Recommended Citation

Li, Xueying C., "Genetic Basis of Thermal Divergence in *Saccharomyces* species" (2018). *Arts & Sciences Electronic Theses and Dissertations*. 1702.

[https://openscholarship.wustl.edu/art\\_sci\\_etds/1702](https://openscholarship.wustl.edu/art_sci_etds/1702)

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact [digital@wumail.wustl.edu](mailto:digital@wumail.wustl.edu).

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences  
Molecular Genetics and Genomics

Dissertation Examination Committee:

Justin C. Fay, Chair  
Yehuda Ben-Shahar  
Barak A. Cohen  
Kenneth Olsen  
Heather L. True-Krob

Genetic Basis of Thermal Divergence in *Saccharomyces* species  
by  
Xueying Li

A dissertation presented to  
The Graduate School  
of Washington University in  
partial fulfillment of the  
requirements for the degree  
of Doctor of Philosophy

December 2018  
St. Louis, Missouri

© 2018, Xueying Li

# Table of Contents

List of Figures ..... viii

List of Supplemental Figures ..... ix

List of Tables ..... x

List of Supplemental Tables ..... x

Acknowledgments..... xi

Abstract..... xvi

Chapter 1: Introduction ..... 1

    Genetic architecture of evolution: from theories to experimental evidence ..... 1

    The role of *cis*-regulatory changes in evolution..... 5

    Thermal divergence in *Saccharomyces* species and the power of yeast genetics ..... 7

    Focus of dissertation work ..... 9

Chapter 2: *Cis*-regulatory divergence in gene expression between two thermally divergent yeast species. .... 12

    Abstract ..... 13

    Introduction ..... 14

    Material and methods ..... 17

        Strains and RNA sequencing ..... 17

        Allele-specific expression differences ..... 17

Quantitative PCR .....	19
Sign test for directional divergence .....	19
Association with genomic features .....	20
Results .....	21
Effects of temperature on allele-specific expression .....	21
Effects of temperature on hybrid gene expression.....	23
Test for temperature-specific directional evolution.....	25
Promoter changes associated with expression divergence .....	27
Discussion .....	29
Environment-dependent cis-effects .....	29
Unexpected heat shock response at low temperatures .....	32
Signatures of selection on cis-acting divergence in gene expression .....	34
Binding sites are only weakly related to expression divergence .....	35
Chapter 3: Mitochondria-encoded genes contribute to evolution of heat and cold tolerance in yeast .....	37
Abstract .....	38
Introduction .....	39
Results .....	41
A non-complementation screen for thermosensitive alleles reveals mitochondrial effects.....	41
Recombinant analysis identifies contribution of multiple mitochondria-encoded genes .....	44
COX1 protein divergence affects both thermotolerance and cryotolerance .....	48

Discussion .....	50
Genetic architecture of interspecies differences in thermotolerance .....	50
Cyto-nuclear interactions in <i>Saccharomyces</i> evolution.....	52
Mitochondrial DNA and yeast evolution.....	55
Materials and Methods .....	55
Strains, growth conditions, and genetic manipulations .....	55
Interspecific hemizygote collections .....	56
Validation of non-complementing genes.....	57
Interspecific hybrids with reciprocal mitotypes .....	58
Crosses with mitochondrial knockouts.....	59
Spontaneous mitochondrial recombinants .....	60
DNA extraction, library preparation, and sequencing .....	60
Mitochondrial genome assembly .....	62
Read mapping and allele assignment of recombinants.....	63
Recombinant phenotypes.....	65
Mitochondrial allele replacement .....	66
Chapter 4: Multiple changes underlie allelic divergence of <i>CUP2</i> between <i>Saccharomyces</i> species.....	68
Abstract .....	69
Introduction .....	70
Materials and Methods .....	71

Results .....	75
Discussion .....	78
Multiple changes with small effects .....	78
Evolution of copper resistance.....	80
Chapter 5: Comparative thermal tolerance of <i>S. cerevisiae</i> and <i>S. uvarum</i> .....	82
Abstract .....	83
Part 1. Correlation between high-temperature growth and heat shock response. ....	84
Introduction .....	84
Materials and Methods .....	85
Results .....	86
Discussion .....	88
Part 2. Effects of lipid supplements on thermotolerance. ....	89
Introduction .....	89
Materials and Methods .....	90
Results and Discussion.....	91
Part 3. Screen for <i>S. cerevisiae</i> genes that enhance <i>S. uvarum</i> 's heat tolerance in <i>S. uvarum</i> using the MoBY-ORF library.....	93
Introduction .....	93
Materials and Methods .....	94
Transformation of MoBY-ORF library and pooling .....	94

Growth competition .....	94
Barcoded sequencing and data analysis .....	95
Validation of plasmid phenotype .....	96
Allele substitution of candidate genes .....	97
GFP strain construction and growth competition .....	101
Integration of MoBY plasmids .....	103
Plasmid burden assay .....	107
Heat shock-based selection on MoBY transformants .....	108
<b>Results .....</b>	<b>109</b>
Screen for <i>S. cerevisiae</i> genes that enhance <i>S. uvarum</i> 's heat tolerance by growth competition .....	109
Validation by individual plasmid transformation .....	110
Allele replacement of candidate genes .....	112
Integration of MoBY plasmids .....	115
Plasmid burden .....	117
Heat shock of MoBY transformants .....	119
<b>Discussion .....</b>	<b>121</b>
Technical challenges in cross-species genetics .....	121
Genetic architecture of heat tolerance .....	123
<b>Chapter 6: Conclusion and future directions .....</b>	<b>125</b>
A complex genetic architecture of interspecific differences .....	125
Mitochondrial genome as an evolutionary hotspot .....	128



Trade-off between heat and cold tolerance .....	130
References .....	134
Appendix .....	161
Supplementary materials for Chapter 2 .....	161
Supplementary materials for Chapter 3 .....	169
Supplementary materials for Chapter 4 .....	185
Supplementary materials for Chapter 5 .....	186

# List of Figures

Figure 2-1: Temperature dependent growth of <i>S. cerevisiae</i> , <i>S. uvarum</i> and their hybrid. ....	16
Figure 2-2: Temperature-dependent allele effects. ....	23
Figure 2-3: Temperature dependent <i>HSP104</i> expression in <i>S. cerevisiae</i> , <i>S. uvarum</i> and their hybrid. ....	25
Figure 3-1: Non-complementation screen for genes underlying phenotypic divergence between <i>S. cerevisiae</i> and <i>S. uvarum</i> . ....	42
Figure 3-2: Mitochondria-encoded genes contribute to divergence in heat and cold tolerance...47	47
Figure 3-3: <i>COX1</i> coding alleles affect growth at high and low temperature. ....	49
Figure 4-1: Design of <i>CUP2</i> chimeras. ....	74
Figure 4-2: Copper resistance of chimeric constructs. ....	76
Figure 5-1: Heat shock survival of <i>S. cerevisiae</i> and <i>S. uvarum</i> . ....	87
Figure 5-2: Trehalose production of <i>S. cerevisiae</i> and <i>S. uvarum</i> at different temperatures.....	88
Figure 5-3: Effects of oleic acid and ergosterol on thermotolerance of <i>S. cerevisiae</i> and <i>S. uvarum</i> . ....	92
Figure 5-4: Scoring scheme for MoBY transformants. ....	97
Figure 5-5: Diagram of <i>ura3</i> alleles constructed in this study. ....	99
Figure 5-6: GFP-tagged <i>S. uvarum</i> strains and their competition with the parental strain. ....	102
Figure 5-7: Diagram of MoBY plasmid. ....	106
Figure 5-8: A positive hit in the growth competition. ....	110
Figure 5-9: Distribution of growth scores of individual MoBY transformants that showed increased heat tolerance compared to the control plasmid. ....	111
Figure 5-10: Spot dilution of MoBY transformants and a non-transformant control. ....	112
Figure 5-11: Allele replacements of 7 candidate genes. ....	113
Figure 5-12: Growth competition with a GFP strain. ....	115
Figure 5-13: Integration of MoBY plasmids. ....	117

## List of Supplemental Figures

Figure S2-1: Transcription factor binding motifs associated with temperature effects. ....	164
Figure S2-2: Correlated binding site and expression changes for <i>ARO9</i> and <i>ARO10</i> . ....	167
Figure S3-1: Reciprocal hemizyosity test of <i>HFA1</i> and <i>CUP2</i> . ....	171
Figure S3-2: Fermentative and respiratory growth of interspecific hybrids with reciprocal mitotypes at different temperatures. ....	172
Figure S3-3: Rescue of <i>S. cerevisiae</i> mitochondrial knockouts by recombination with <i>S. uvarum</i> mitotypes. ....	173
Figure S3-4: Recombinant genotypes and examples of recombination breakpoints. ....	174
Figure S3-5: High petite rate of <i>S. uvarum</i> mitotype and its association with <i>ORF1</i> . ....	177
Figure S3-6: Procedure for mitochondrial allele replacement. ....	179
Figure S3-7: Background-dependent allele effects of <i>COX1</i> . ....	181
Figure S4-1: Copper resistance of chimeric constructs in <i>S. cerevisiae</i> . ....	185

## List of Tables

Table 2-1: Groups of genes showing directional evolution at three temperatures.....	26
Table 4-1: Additive and epistatic effects of <i>S. cerevisiae</i> <i>CUP2</i> regions on copper resistance. .	77
Table 5-1: Primers used in MoBY integration. ....	107
Table 5-2: Plasmid loss per generation in MoBY transformants. ....	118
Table 5-3: Heat shock survival rate after 5 HS-recovery cycles. ....	120
Table 5-4: HS survival of individual colonies in the HS pool. ....	121

## List of Supplemental Tables

Table S2-1: GO terms enriched in temperature-responsive genes. ....	161
Table S3-1: Aneuploidy in the recombinants. ....	182
Table S3-2: Strains used in this study. ....	183

# Acknowledgments

I am sincerely grateful to everyone I met during the PhD journey. All of them make me know better what kind of person I want to be. I especially thank my advisor, Justin Fay, to supervise the research presented in this dissertation. I thank him for his brilliant ideas and constant inspiration in the science problems, as well as the support and training in every aspect of my research life, from little things like correcting my grammatical errors and fixing equipment, to opportunities of science conferences, mentoring undergrads and peer reviews. He is a role model both in his consistently high-quality work and his care for others.

I thank my committee members for their support during the last five years. They have provided constructive input for my research and helped me navigate the PhD and be practical about the timeline. They encouraged me to pursue the mitochondrial study, which turned out to be the major body of this dissertation work. Yehuda Ben-Shahar mentored me during my rotation, and he always inspires me with his imagination and insight on biology and evolution. Barak Cohen deeply influenced me with his philosophy of doing science and his poetic character; both Barak and Heather True provided valuable suggestions with their expertise in yeast genetics. Ken Olsen teaches the Evolution class for which I was a teaching assistant, and I learned a lot from his work ethics and his insight on evolutionary genetics.

I thank my collaborators, David Peris and Chris Todd Hittinger at University of Wisconsin-Madison (Peris is now at Institute of Agrochemistry and Food Technology in Spain), and Elaine Sia at University of Rochester. Peris and Chris provided the mitochondrial genome assembly of *S. uvarum* and helped revise the manuscript reprinted in Chapter 3. Elaine Sia helped me design the mitochondrial allele replacements and taught me the technique of biolistic transformation. I also thank Tom D. Fox, Alexander Tzagoloff, Javier Alonso del Real Arias and

Amparo Querol for sharing strains. Additionally, I thank National Institutes of Health for funding the research.

I thank all the past and current members in Fay lab for a very supportive lab environment. Ping Liu, Andrew Bergen, Kim Lorenz and Linda Riles helped me get started with yeast techniques and offered generous advice when I met problems in wet lab experiments. Ping Liu and Kim Lorenz particularly contributed to the MoBY screen and plasmid validation, respectively, in Chapter 5. In my early years of the PhD, Kim as a postdoc and an MGG-alumni showed me how to think and work as a mature researcher. Linda and I used to be “baymates” and she showed me how young a person in her seventies can be. Ching-hua Shih helped with the RNA-seq analysis pipeline and taught me basic Linux commands. He was also my labmate for the longest time; we went through lab moves between buildings and between cities, and there were many things that I thought I could do alone, only to realize I couldn’t without his help. I also thank Emery Longan for sharing the lab move experience and for being an important source of input of all the research discussions. All of them, along with Katie Williams, Brett Tortelli, and Nilima Walunjkar, are good friends that I could share any thought with, no matter about science or life. I thank all my undergrad mentees, Camille Rieber, Haley Cohen and Sabrina Sun, who trained me to be a better mentor. Camille also assisted in some phenotypic assays in Chapter 3.

I thank the MGG program directors, Jim Skeath, Tim Schedl and John Edwards, and the student coordinator Melanie Relich for running literally the best Genetics graduate program. Jim influenced me greatly with his passion in teaching and education and helped me improve my presentation skills. I thank the DBBS student community for being great peers. From seminars

and classes, I learned about their research and ideas, many of which are eye-opening. They are such a group of friendly and intelligent people, and I am grateful getting to know them.

I thank the CGS community, particularly Stormo lab, Cohen lab, Brent lab and Gordon lab, for sharing space, reagents and equipment. I thank Devi Swain Lenz for advice in yeast techniques and Ashley Wolf for help with qPCR. I thank Jess Hoisington-Lopez for her excellence in running the sequencing core; only after I left CGS did I realize how indispensable she is. CGS is such a supportive environment where I know I can always find help, and the research discussions there were truly vibrant and inspiring.

I thank Department of Biology at University of Rochester, where I made friends with shared scientific interests, and deepened my understanding in many evolutionary genetics questions. I thank Larracuenta lab, Lambert lab, Sia lab and Fu lab for sharing equipment, particularly David Lambert for help with fluorescent microscopes. I thank Cindy Landry for help with my enrollment as a visiting student and Mary Bissell for help during the lab move.

I thank my parents for supporting my study in a foreign country; I wish I could have spent more time with them during the past years. I thank all my friends for their care and support, no matter in St. Louis, Rochester, or other parts of the world. I thank my roommates at the 5793 house, both humans and animals, who made the two-year time a remarkable era: I thank Cheng Cheng for broadening my world with her “extreme tourism” and Qihao Ren for his generosity and humor. I thank the cat Cat, Latte, Toothless and Sophie, the chameleon Neuvo and two anonymous birds for their unpretentious behavior and for nourishing my curiosity about life and nature. I particularly thank cat Latte for taking care of me at Rochester. Finally, I thank Xitong Liang, for being my best friend, my mentor, and my source of inspiration and happiness. His passion and curiosity in biology drives me to hold on to science, and he works so hard that I feel

I should do the same. He cheered me up a million times in the face of frustration and struggles, with his love and inspiration. This PhD would not be possible without him.

Xueying Li

*Washington University in St. Louis*

*December 2018*



Dedicated to *life*.

## ABSTRACT OF THE DISSERTATION

Genetic Basis of Thermal Divergence in *Saccharomyces* species

by

Xueying Li

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics and Genomics

Washington University in St. Louis, 2018

Professor Justin C. Fay, Chair

The genetic architecture of phenotypic divergence is a central question in evolutionary biology. Genetic architecture is impacted by whether evolution occurs through accumulation of many small-effect or a few large-effect changes, the relative contribution of coding and *cis*-regulatory changes, and the prevalence of epistatic effects. Our empirical understanding of the genetic basis of evolutionary change remains incomplete, largely because reproductive barriers limit genetic analysis to those phenotypes that distinguish closely related species. In this dissertation, I use hybrid genetic analysis to examine the basis of thermal divergence between two post-zygotically isolated species, *Saccharomyces cerevisiae* and *S. uvarum*. *S. cerevisiae* is relatively heat tolerant, whereas *S. uvarum* is heat sensitive but outperforms *S. cerevisiae* at 4°C. Gene expression analysis with an *S. cerevisiae* and *S. uvarum* hybrid revealed a small set of 136 genes with temperature-dependent *cis*-acting differences, suggesting that the temperature divergence has not caused widespread *cis*-regulatory divergence. Using a genome-wide non-complementation screen, I found a single nuclear-encoded gene with a modest contribution to heat tolerance, and a large effect of the species' mitochondrial DNA (mitotype). Recombinant

mitotypes and allele replacements indicate multiple mitochondria-encoded genes contribute to thermal divergence, with the coding sequence of *COX1* showing a moderate effect on both heat and cold tolerance. The non-complementation approach also identified allele differences of *CUP2*, a copper-binding transcription factor, in copper resistance of *S. cerevisiae* and *S. uvarum*. Chimeric alleles showed that multiple changes underlie the resistance of *S. cerevisiae* *CUP2*, with *cis*-regulatory changes having a larger effect than coding changes. Taken together, my findings suggest that evolution of interspecific phenotypic differences often involves accumulation of small-to-medium effect changes, such as those in mitochondrial DNA and *CUP2*, and can occur through both coding and *cis*-regulatory changes.

# Chapter 1: Introduction

The genetic architecture of phenotypic divergence is a central question in evolutionary genetics. The genetic basis of species' difference can be substantially different from that of phenotypic variation within species. Over a long evolutionary timescale, species may accumulate many small-effect changes that contribute to their adaptive phenotypes, as suggested by the model of micromutationism. However, the micromutationist model has been debated and it remains unresolved how many changes underlie evolution and what is the distribution of their effect sizes. Empirical studies have found a mixture of small- and large-effects, but bias in experimental approaches could have shaped our view of evolution. In addition, *cis*-regulatory changes may play an important role in phenotypic evolution, but the relative contribution of *cis*-regulatory and coding changes is unclear. Additionally, the prevalence of epistasis in species' phenotypic divergence remains largely unknown. In this chapter, I will review our current understanding regarding these questions and the challenges the field has met. I will introduce my study system, thermal divergence in yeast species, and how high-throughput experiments with yeast would address the evolutionary questions in a systematic, relatively unbiased manner.

## **Genetic architecture of evolution: from theories to experimental evidence**

There has been a long-standing debate about the number of changes underlie evolution and the distribution of their effect sizes. The model of micromutationism originated from Darwin (Darwin 1859), positing that evolution occurs primarily through accumulation of many small-effect changes, whereas the alternative view emphasizes the role of large-effect changes in evolution. During the modern synthesis, it was generally accepted that many small-effect

changes contribute to evolution. Fisher proposed the geometric model of adaptation and showed that mutations of small effects are more likely to be favorable than those with large effects (Fisher 1930). Thus, mutations underlying adaptation are likely to have infinitesimally small effects. However, little empirical evidence was found for this micromutationism view, since many quantitative trait locus (QTL) studies found large-effect loci underlying evolution (Orr and Coyne 1992). As a further development of Fisher's geometric model, Orr showed that the effect size of factors fixed during adaptation follows an exponential distribution, including a few large-effect changes and many small-effect changes (Orr 1998). This model is in better agreement with empirical findings and has been generally accepted. For example, a recent review of 28 published QTL studies supported an exponential distribution of effect size of genetic factors underlying adaptation of threespine stickleback (Peichel and Marques 2016).

To summarize the empirical evidence and test hypotheses about the genetic architecture, multiple authors have compiled lists of genetic variants identified in past experimental studies (Orr 2001; Hoekstra and Coyne 2007; Stern and Orgogozo 2008; Martin and Orgogozo 2013). From these reviews, there appears no universal genetic architecture of evolution: both small- and large-effect loci have been found, and the traits can be either Mendelian or polygenic. For example, studies of three different traits that diverged in *Drosophila* species revealed three different types of genetic architecture: divergence in larval hairs was explained by a single gene (Sucena and Stern 2000), evolution of adult toxin resistance involved more than five genes (Jones 1998), and the difference in posterior lobe size/shape was mapped to more than 19 QTLs (Zeng et al. 2000) (reviewed by (Orr 2001)). This heterogenous pattern might not be surprising, since each trait may have a unique evolutionary trajectory. However, with the accumulation of

more experimental evidence, we might be able to stratify the data by evolutionary divergence, strength of selection, and other factors, to derive generalized principles.

In addition to the number of changes and the distribution of effect sizes, the prevalence of epistasis is also relevant to the genetic architecture of evolution. Epistasis is important to evolution because the identity of later mutations is constrained by the mutations fixed earlier. However, epistatic interactions between phenotypically relevant loci were more often observed in intraspecific variations than interspecific comparisons (Orr 2001; Stern and Orgogozo 2009). Compared to the epistatic alleles whose effects depend on the genetic background, selection might favor non-epistatic alleles that have the same effect (e.g. increase growth) in all backgrounds (Stern and Orgogozo 2009). However, interpretation of epistasis in experimental hybrids may require caution because the epistatic mutations might not be co-selected in the parent species. Mutations might be fixed sequentially in the parental species, in which case they are selected solely for their additive effects. Therefore, the contribution of epistasis to adaptation cannot be interpreted with high confidence from these experimental studies, compared to additive effects (Orr 2001). Because of the problem in interpretation, the role of epistasis in interspecific differences remains inconclusive.

Despite recent progress, our empirical understanding of evolution could be biased for several reasons. First, identification of large-effect loci may inevitably suffer from detection bias (Rockman 2011). Large-effect loci are by definition easier to identify than small effects, and one may argue that the statistical power of many QTL study designs only allowed detection of large effects (Orr 2001; Rockman 2011). Failure to detect large effects is often regarded as a negative result and unlikely to be published. Secondly, large-effect loci or genes may represent aggregation of many small-effect nucleotide changes, but few studies have solved the causal

changes to single nucleotide resolution. Indeed, when high-resolution mapping was performed with *shavenbaby*, a causal gene for evolution of naked cuticle in *Drosophila* larvae, its large-effect was found to be caused by many subtle-effect substitutions (Frankel et al. 2011). Such “hotspot” locus reconciles the micromutationist theories with the empirical observation of large-effect QTLs (Stern and Orgogozo 2009; Martin and Orgogozo 2013), and also highlights the need for genetic dissection at single-nucleotide resolution.

Bias could also result from the approaches and subjects available for interspecific genetic analysis. QTL mapping allows a genome-wide survey for causal loci, but it can only be applied to closely related species that are interfertile. The taxa mostly analyzed by QTL approaches include the *Drosophila simulans* complex and *Mimulus guttatus* (monkeyflower) complex, both diverged less than 1 million years ago (Mya) (Beardsley et al. 2004; Tamura et al. 2004; Brandvain et al. 2014). The genetic architecture of recent divergence might not be fully representative of adaptation in a longer timescale. For example, QTL studies of sticklebacks have revealed hotspot genes underlying adaptation to freshwater environment, which happened less than 12,000 years ago when marine populations colonized freshwater lakes (Schluter et al. 2010; Peichel and Marques 2016). The strong selection of salt-to-freshwater transition during a short time might have favored certain large-effect loci. Large-effect QTLs were also often found to underlie domestication (Purugganan and Fuller 2009; Wright 2015) and might be associated with strong selection, although recent evidence showed that the genetic architecture of domestication might be more complex than previously envisioned, and the strength of selection may be variable (reviewed in (Purugganan and Fuller 2009)).

Distantly related species cannot be analyzed by the QTL approach because of reproductive barriers. Comparative studies of reproductively isolated species often rely on

candidate approaches, which naturally lead to a focus on single genes of large effect. For example, the *yellow* gene is required for the formation of black melanin in *Drosophila melanogaster* (fruitfly) and was selected as a candidate for evolution of pigmentation in *Drosophila* species. Indeed, expression divergence of *yellow* correlated with divergent melanin patterns in *Drosophila* species that span a genetic distance of 65 Mya (Wittkopp et al. 2002), and a series of studies subsequently revealed how this pathway diversified within and between species (Massey and Wittkopp 2016). Such detailed dissection of developmental genes has provided important insight on the molecular mechanisms of evolution, specifically that *cis*-regulatory changes often underlie morphological evolution between species (Carroll 2000) (see below). Despite its high resolution, the candidate approach might not provide a comprehensive view of the number of genes involved in divergence, leaving our knowledge of phenotypic divergence between distantly related species incomplete.

Taken together, our empirical view of evolution could be shaped by the approaches used. Our experiments were inevitably biased towards detection of large-effect changes, and reproductive barriers prevented genome-wide surveys of distantly related species. These gaps could be potentially filled by improved study designs, with higher resolution and genome-wide screens. In this regard, high-throughput experiments with divergent yeast species may be a good starting point, as will be presented in this dissertation.

### **The role of *cis*-regulatory changes in evolution**

Comparative studies of morphological traits have given rise to the “*cis*-regulatory hypothesis”, stating that most mutations causing morphological variation are in the *cis*-regulatory, rather than the coding, regions of developmental genes (Stern 2000). The theory is



based on a modular nature of development: developmental genes are often regulated by multiple *cis*-regulatory elements (CREs), with each CRE as an independent module that turns on/off the target gene in a specific tissue or at a specific time. Compared to coding mutations, *cis*-regulatory mutations are less likely to be pleiotropic and thus more likely to be favored by natural selection. This hypothesis has been debated. Through analysis of published genetic variants underlying phenotypic differences, Hoekstra and Coyne found that most adaptive phenotypes involved changes in protein coding sequences (Hoekstra and Coyne 2007). Stern and Orgogozo argued that, although fewer *cis*-regulatory variants had been uncovered than coding variants (presumably due to the technical difficulties in mapping regulatory variants), they were enriched in morphological evolution and evolution above the species level (Stern and Orgogozo 2008). Although the relative contribution of *cis*-regulatory and coding mutations might be debatable, it is now generally accepted that *cis*-regulatory changes can play an important role in phenotypic diversity within and between species (Wittkopp and Kalay 2012).

In addition to the molecular studies reviewed by the authors above, *cis*-regulatory changes have also been studied for their effect on gene expression. Gene expression is critical to biological processes and is subject to natural selection (Fay and Wittkopp 2008). By analyzing allele-specific expression in F1 hybrids, many studies have shown that *cis*-regulatory divergence has caused widespread gene expression differences between species. *Trans*-acting changes also contribute to expression divergence, and their prevalence varies among cases (Wittkopp et al. 2004; Tirosh et al. 2009; McManus et al. 2010). In several cases, *cis*-regulatory mutations in functionally related or co-regulated genes were found to change the expression levels in the same direction (e.g. consistently higher in one species than the other), suggesting directional evolution of these pathways (Bullard et al. 2010; Fraser et al. 2012). Despite the widespread *cis*-effects, the

*cis*-regulatory variants underlying gene expression divergence have been difficult to pinpoint. On a genome-wide scale, expression divergence does not seem to correlate with differences in transcription factor binding sites (Tirosch et al. 2008), coding sequences (Tirosch and Barkai 2008), or nucleosome positioning (Tirosch et al. 2010). The lack of genome-wide pattern suggests that understanding expression divergence of individual genes might require complex empirical work (Wittkopp and Kalay 2012).

While *cis*-regulatory evolution is known to affect gene expression, the link between the level of gene expression and organismal phenotypes is not always straightforward. In the case of evo-devo studies, expression patterns are often associated with cell fate and thus their phenotypic consequences are relatively clear (Stern and Orgogozo 2008). For non-developmental traits, gene expression was correlated with phenotypic variation in a few cases, often in a small number of genes (Fay et al. 2004; Engle and Fay 2012), but did not predict fitness in a linear fashion in other cases (Bergen et al. 2016; Keren et al. 2016; Duveau et al. 2017). These findings highlight the complexity in understanding the role of *cis*-regulatory changes in phenotypic evolution.

### **Thermal divergence in *Saccharomyces* species and the power of yeast genetics**

The *Saccharomyces* group is a rising model for evolutionary genetics studies (Scannell et al. 2011; Caudy et al. 2013). *S. cerevisiae* as a model organism in classical genetics has been extensively studied for almost a century (Barnett 2007). *S. paradoxus*, *S. mikatae*, *S. kudriavzevii* and *S. uvarum* diverge from *S. cerevisiae* at different genetic distances, with *S. cerevisiae* and *S. uvarum* (the basal species in the lineage) being more different at synonymous sites than human and mouse (Waterston et al. 2002; Kawahara and Imanishi 2007). The *Saccharomyces* species diverged after the whole genome duplication in the budding yeast lineage and have a

fermentative lifestyle (preference for fermentation in the presence of oxygen), a trait associated with adaptation to high-sugar environment. While *Saccharomyces* are widely found in arboreal habitats such as oak trees, *S. cerevisiae* is often present in vineyards and many human-associated fermentations, including beer and wine production and baking. *S. kudriavzevii*, *S. uvarum*, *S. eubayanus* and their hybrids with *S. cerevisiae* are used in low-temperature fermentations, such as industrial lager or cider brewing (reviewed in (Dashko et al. 2016)).

Perhaps the largest phenotypic difference among the *Saccharomyces* species is their thermal growth profile (Gonçalves et al. 2011; Salvadó et al. 2011). Both *S. cerevisiae* and *S. paradoxus* can grow at high temperature (37°C), despite a difference in the upper limit (42°C for *S. cerevisiae* and 38°C for *S. paradoxus*) (Gonçalves et al. 2011). The more distantly related species *S. kudriavzevii* and *S. uvarum* are cold tolerant, thriving at lower than 10°C and unable to grow above 34°C. The genetic basis of the thermal divergence among these species is largely unclear. Previous analysis identified a small number of candidate genes, such as glycolytic enzymes (Gonçalves et al. 2011) and genes involved in lipid metabolism, oxidoreductase and vitamin pathways (Paget et al. 2014). Only one systematic screen has been performed, which found eight housekeeping genes underlying the difference in 39°C-growth between *S. cerevisiae* and *S. paradoxus* (Weiss et al. 2018). This finding suggests that thermal divergence between *S. cerevisiae* and the cryotolerant species *S. uvarum* may involve many more genetic changes, potentially including changes in the fundamental growth machinery of yeast cells.

The growth temperature divergence between *S. cerevisiae* and *S. uvarum* provides an experimental system for tackling the genetic architecture of phenotypic divergence. The *Saccharomyces* species are post-zygotically isolated: haploids of different species are able to

form F1 diploid hybrids, but the hybrids cannot produce viable spores. The F1 hybrids of *S. cerevisiae* and *S. uvarum* are both heat- and cold-tolerant, providing an opportunity to examine the two diverged genomes in the same *trans*-environment. The ease of high-throughput genetic manipulations in yeast enables mapping interspecific differences in a systematic, high-resolution manner. The yeast deletion collection was released in 2000, allowing genome-wide screens for knockout phenotypes or non-complementation. The recent advances in next-generation sequencing and genome editing techniques have made it possible to measure phenotypic effects of thousands of genotypes in massive parallel. For example, different genotypes can be labelled with oligonucleotide “barcodes” and grown in a pool, and their fitness can be determined by the abundance of barcodes quantified by sequencing (Ho et al. 2009; Levy et al. 2015). Fitness effects of single nucleotide variants can also be characterized in a high-throughput manner (Metzger et al. 2015; Li et al. 2016; Sharon et al. 2018), providing possibilities for testing evolutionary hypotheses at a high resolution. Relevant to mapping genetic variants underlying phenotypic divergence, methods in F1 hybrids like CRISPR-induced mitotic recombination (Sadhu et al. 2016) and reciprocal hemizyosity analysis via sequencing (RH-seq) (Weiss et al. 2018) have overcome the problem of reproductive barriers and are expected to provide new sights for evolution of species differences.

### **Focus of dissertation work**

A comprehensive understanding of genetic architecture of evolution may require systematic approaches, to avoid potential bias in candidate selection or mapping design. Thermal divergence between two post-zygotically isolated yeast species, *S. cerevisiae* and *S. uvarum*, provides a good experimental platform for such approaches. In this dissertation, I present my

work using genome-wide screens to identify genes underlying thermal divergence between *S. cerevisiae* and *S. uvarum*. By mating the *S. cerevisiae* deletion collection to *S. uvarum* and screening for heat sensitive hemizygotes, I was able to systematically assess allele effects of more than 4,700 non-essential genes. A complementary screen was performed by introducing a library of more than 4,000 *S. cerevisiae* genes into *S. uvarum* and testing their phenotypic effects via barcode sequencing, although this approach has met technical complications. For the large-effect genes that came out from the screen, I tested whether their effects are caused by multiple small-effect changes, and the relative contribution of *cis*-regulatory and coding changes. This includes fine-mapping within the mitochondrial genome for its temperature effect. I am also interested in the effects of *cis*-regulatory divergence on gene expression, which was addressed by RNA sequencing.

These efforts will be presented in Chapter 2-5. In Chapter 2, I studied gene expression of a hybrid of *S. cerevisiae* and *S. uvarum* at different temperatures. Using a previously developed sign test, I tested for signatures of directional selection in relation to temperature response. In Chapter 3, I performed a genome-wide non-complementation screen for genes underlying heat tolerance between the two species. The screen revealed a single nuclear-encoded gene (*HFAI*) with a modest effect, and a large-effect of mitochondrial DNA. The latter was further dissected via recombination mapping and allele replacements, to identify mitochondria-encoded genes involved in thermal divergence and their effect sizes. In Chapter 4, I analyzed *CUP2*, a gene showing non-complementation for copper resistance, to test whether multiple changes underlie *CUP2* divergence and to examine the effects of *cis*-regulatory and coding changes. Chapter 5 describes a second genome-wide screen using a library of *S. cerevisiae* genes, including the

complication of plasmid burden and how it might be solved by studying survival at lethal heat shock temperatures.

The results collectively support a polygenic architecture underlying phenotypic divergence between species: multiple mitochondrial-encoded genes contributed to the evolution of heat and cold tolerance in *Saccharomyces* species; *CUP2* showed a moderate allele difference in copper resistance, and its effect was caused by multiple changes in both coding and *cis*-regulatory regions, with the proximal promoter showing the largest effect. *Cis*-regulatory evolution had prevalent effects on gene expression divergence but did not cause wide-spread heat sensitivity in gene expression. Although having not resolved every genetic change underlying thermal divergence, my research represents one of the few genome-wide screens so far conducted for mapping interspecific differences. The mitochondrial effects on thermal divergence found in this study, along with prior reports of mito-nuclear incompatibilities in reproductive isolation (Lee et al. 2008; Chou et al. 2010), point to the mitochondrial genome as an evolutionary hotspot in yeast speciation and adaptation. The findings together highlight a complex genetic architecture of interspecific differences, showing that evolution may occur through multiple small-to-moderate effect changes, and both coding and *cis*-regulatory.

## **Chapter 2: *Cis*-regulatory divergence in gene expression between two thermally divergent yeast species.**

This chapter was done in collaboration with Justin C. Fay and is a reprint from a 2017 publication in *Genome Biology and Evolution*.

## Abstract

Gene regulation is a ubiquitous mechanism by which organisms respond to their environment. While organisms are often found to be adapted to the environments they experience, the role of gene regulation in environmental adaptation is not often known. In this study, we examine divergence in *cis*-regulatory effects between two *Saccharomyces* species, *S. cerevisiae* and *S. uvarum*, that have substantially diverged in their thermal growth profile. We measured allele specific expression (ASE) in the species' hybrid at three temperatures, the highest of which is lethal to *S. uvarum* but not the hybrid or *S. cerevisiae*. We find that *S. uvarum* alleles can be expressed at the same level as *S. cerevisiae* alleles at high temperature and most *cis*-acting differences in gene expression are not dependent on temperature. While a small set of 136 genes show temperature-dependent ASE, we find no indication that signatures of directional *cis*-regulatory evolution are associated with temperature. Within promoter regions we find binding sites enriched upstream of temperature responsive genes, but only weak correlations between binding site and expression divergence. Our results indicate that temperature divergence between *S. cerevisiae* and *S. uvarum* has not caused widespread divergence in *cis*-regulatory activity, but point to a small subset of genes where the species' alleles show differences in magnitude or opposite responses to temperature. The difficulty of explaining divergence in *cis*-regulatory sequences with models of transcription factor binding sites and nucleosome positioning highlights the importance of identifying mutations that underlie *cis*-regulatory divergence between species.



## Introduction

Changes in gene regulation are thought to play an important role in evolution (Carroll 2000). Regulatory change may be of particular importance to morphological evolution where tissue specific changes and co-option of existing pathways can modulate essential and conserved developmental pathways without a cost imposed by more pleiotropic changes in protein structure. Indeed, many examples illustrate this view and there is a strong tendency for *cis*-acting changes in gene expression to underlie morphological evolution between species (Stern and Orgogozo 2008).

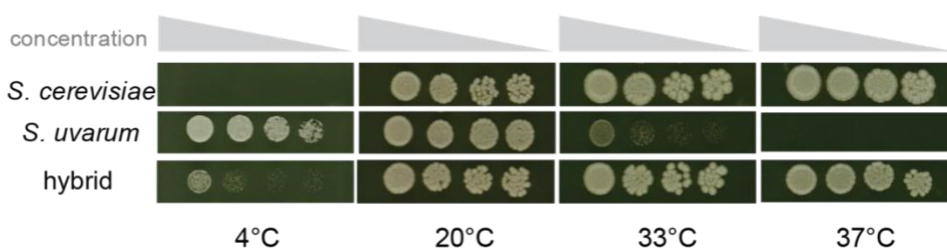
However, gene regulation is also critical to responding to environmental changes and all organisms that have been examined exhibit diverse transcriptional responses that depend on the environmental alteration (López-maury et al. 2008). Environment-dependent gene regulation enables fine-tuning of metabolism depending on nutrient availability as well as avoiding the potential costs of constitutive expression of proteins that are beneficial in certain environments but deleterious in others. Despite the general importance of responding to changing environments, the role of gene regulation in modulating these responses between closely related species is not known and may involve structural changes in proteins whose expression is already environment-dependent.

Studies of genetic variation in gene expression within and between species have revealed an abundance of variation (reviewed in (Whitehead and Crawford 2006; Zheng et al. 2011; Romero et al. 2012)). When examined, a significant fraction of this variation is environment-dependent ((Fay et al. 2004; Landry et al. 2006; Li et al. 2006; Smith and Kruglyak 2008; Tirosh et al. 2009; Fear et al. 2016; He et al. 2016); reviewed in (Gibson 2008; Grishkevich and Yanai

2013)). However, distinguishing between adaptive and neutral divergence in gene expression is challenging (Fay and Wittkopp 2008), since *trans*-acting changes can cause correlated changes in the expression of many genes and the rate of expression divergence depends on the mutation rate and effect size, which is likely gene-specific and not known for all but a few genes (Gruber et al. 2012; Yun et al. 2012; Metzger et al. 2015).

One potentially powerful means of identifying adaptive divergence in gene expression is through a sign test of directional *cis*-acting changes in gene expression measured by allele-specific expression (ASE) (Fraser 2011). By testing whether a group of functionally related or co-regulated group of genes have evolved consistently higher or lower expression levels, the test does not assume any distribution of effect sizes and more importantly is specifically targeted to identifying polygenic adaptation. Applications of this or related sign tests (Fraser et al. 2010; Naranjo et al. 2015) have revealed quite a few cases of adaptive evolution (Bullard et al. 2010; Fraser et al. 2010; Fraser et al. 2011; Fraser et al. 2012; Martin et al. 2012; J. Chang et al. 2013; Naranjo et al. 2015; He et al. 2016; Roop et al. 2016), some of which have been linked to organismal phenotypes. However, in only two of these studies was condition-specific divergence in gene expression examined (He et al. 2016; Roop et al. 2016), leaving open the question of how often such changes exhibit evidence for adaptive evolution. Of potential relevance, the majority (44-89%) of environment-dependent differences in gene expression have been found to be caused by *trans*- rather than *cis*-acting changes in gene expression (Smith and Kruglyak 2008; Tirosh et al. 2009; Grundberg et al. 2011; Fear et al. 2016), suggesting that *trans*-acting changes in gene expression may be more important to modulating environment-dependent gene expression.

In this study, we examine allele-specific differences in expression between two *Saccharomyces* species that have diverged in their thermal growth profiles. Among the *Saccharomyces* species, the most prominent phenotypic difference is in their thermal growth profile (Gonçalves et al. 2011; Salvadó et al. 2011). The optimum growth temperature of *S. cerevisiae* and *S. paradoxus* is 29-35°C, whereas the optimum growth temperature for *S. uvarum* and *S. kudriavzevii* is 23-27°C (Salvadó et al. 2011). Furthermore, *S. cerevisiae* is able to grow at much higher temperatures (maximum 41-42°C) than *S. uvarum* (maximum 34-35°C, (Gonçalves et al. 2011)), while *S. uvarum* grows much better than *S. cerevisiae* at low temperature (4°C, Fig. 2-1). Because *S. cerevisiae* × *S. uvarum* hybrids grow well at high temperature, we were able to measure *cis*-regulatory divergence in gene expression across a range of temperatures by measuring ASE in the hybrid. We use this approach to determine how ASE is influenced by temperature and specifically whether *S. uvarum* alleles are misregulated at temperatures not experienced in their native context. We find that most ASE is independent of temperature and only a small subset of genes show an allele-specific temperature response.



**Figure 2-1. Temperature dependent growth of *S. cerevisiae*, *S. uvarum* and their hybrid.**

Growth is after 17 days at 4°C, 3 days at 20°C and 2 days at 33°C and 37°C, with platings on YPD at 1:3 serial dilutions.

## Material and methods

### *Strains and RNA sequencing*

A hybrid strain YJF1484 was made by crossing an *S. cerevisiae* strain YJF153 (*MATa hoΔ::dsdAMX4*, derived from an oak tree isolate YPS163) and an *S. uvarum* strain YJF1450 (*MATa hoΔ::NatMX*, derived from CBS7001 and provided by C. Hittinger). The hybrid was typed by PCR (Albertin et al. 2013) and found to carry *S. cerevisiae* mitochondrial DNA. A diploid *S. cerevisiae* strain YJF1463 was made by crossing YJF153 (*MATa hoΔ::dsdAMX4*) and YJF154 (*MATa hoΔ::dsdAMX4*, derived from YPS163). The diploid *S. uvarum* strain YJF2602 was made by crossing YJF1449 (*MATa, hoΔ::NatMX*, derived from CBS7001) and YJF1450 (*MATa hoΔ::NatMX*).

Three replicate overnight cultures of the diploid hybrid YJF1484 were used to inoculate 50 ml YPD cultures (1% yeast extract, 2% peptone, 2% glucose) and incubated at either 22°C, 33°C or 37°C at 300 rpm. Cells were harvested at mid-log phase and RNA was extracted with phenol/chloroform. The nine RNA samples were enriched for mRNA by poly A purification, reverse transcribed, fragmented, ligated to indexed adaptors and sequenced on a HiSeq (1×50 bp run) at Washington University's Genome Technology Access Center. The RNA-seq reads were deposited at Sequence Read Archive under the accession number SRP168939.

### *Allele-specific expression differences*

Reads were mapped using Bowtie2 (Langmead and Salzberg 2012) to a combined *S. cerevisiae* and *S. uvarum* genome. The YJF153 genome was generated by converting the S288c (R64-1-1) reference to YJF153 using GATK (v3.3-0) and YJF153 variants. YJF153 variants

were called using GATK and 5.3 million paired-end (2×101 bp) HiSeq reads (SRX2321838). Annotations for the YJF153 genome were obtained using S288c annotations and the UCSC LiftOver tool. The YJF1450 genome and annotation files were obtained from Scannell et al. (2011). We obtained an average of 5.5 million mapped reads per sample after removing duplicate reads and reads with low mapping quality (MQ < 2). All the remaining reads were uniquely mapped as they had a higher primary than secondary alignment score (AS>XS). Read counts for each gene were generated using HTSeq-count (Anders et al. 2015) with the default settings, which only counts reads with a mapping quality of at least 10. Species-specific counts of 5,055 orthologs were generated using previously defined one-to-one orthologs (Scannell et al. 2011). To quantify any systematic bias in read mapping we calculated the ratio of normalized *S. cerevisiae* to *S. uvarum* expression levels and found a median of 0.998, indicating no systematic read mapping bias. In our data, expression differences did not correlate with GC content (p = 0.74, linear regression), which was a concern in a previous report (Bullard et al. 2010).

Significant differences in expression were tested using a generalized linear model with a negative binomial error model (Anders et al. 2010). Using normalized read counts we tested each gene for i) temperature effects, ii) allele effects, and iii) temperature-allele interactions by dropping terms from the full model:  $counts \sim allele + temperature + allele*temperature$ , where *allele* and *temperature* are terms indicating the species' allele and temperature effect and the star indicates an interaction. For score assignment in the sign test (see below), we treated data from three temperatures separately and tested each gene for allele-specific expression (ASE) at each temperature. A false discovery rate (FDR) cutoff of 0.05 was used for significance.

### *Quantitative PCR*

Quantitative PCR (qPCR) was used to quantify the expression of *HSP104* in the hybrid as well as both parental strains following temperature treatment. Overnight cultures were grown at 23°C, diluted to an optical density (OD<sub>600</sub>) of 0.1 in YPD for temperature treatment and grown at 10°C, 23°C and 37°C for two days, 6 hours and 5 hours respectively. The middle and high temperature cultures were shaken at 250 rpm whereas the low temperature cultures were grown without shaking. At the time of collection, the OD<sub>600</sub> of the cultures were all within the range of 0.5 - 1.9. RNA was extracted as described above, DNase I treated (RQ1 RNase-free DNase, Promega) and cDNA was synthesized (Protoscript II Reverse Transcriptase, New England Biolabs). qPCR amplifications used the Power SYBR Green Master Mix (Thermo Fisher Scientific Inc.) and were quantified on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Each PCR reaction was run in triplicate and one sample was removed from analysis due to a high standard error of deltaCt values (>0.4) among the three technical replicates. For each sample, expression of *HSP104* was measured relative to *ACT1* expression. Because we used allele-specific primers to distinguish *S. cerevisiae* and *S. uvarum* alleles of *HSP104*, the expression levels were corrected using the PCR efficiency of each primer sets, determined by standard curves. Genomic DNA of YJF1484 was used as a calibrator and to remove any plate-to-plate differences.

### *Sign test for directional divergence*

Pathways and groups of co-regulated genes were tested for directional divergence using a sign test as previously described (Bullard et al. 2010). Each gene was assigned a score 0 if the gene showed no ASE, 1 if the gene showed ASE and the *S. cerevisiae* allele was expressed

higher than the *S. uvarum* allele and -1 if the gene showed ASE and the *S. cerevisiae* allele was expressed lower than the *S. uvarum* allele. Scores for all the genes in a co-regulated group (Gasch et al. 2004) were summed and tested for significant deviations from 0 by permutation resampling of scores across all 5055 genes. To correct for multiple comparisons, the false discovery rate was estimated from the permuted data across all groups. The analysis was independently applied to data from 22°C, 33°C and 37°C.

#### *Association with genomic features*

Expression levels were associated with features of intergenic sequences, defined as sequences between annotated coding sequences. Intergenic sequences were obtained from <http://www.SaccharomycesSensuStricto.org> and pairwise alignments were generated using FSA (Bradley et al. 2009). Substitution rates were calculated using the HKY85 model of nucleotide substitution implemented in PAML (Yang 2007).

Transcription factor binding site scores were generated by Patser (Hertz and Stormo 1999) with 244 position weight matrix (PWM) models from YeTFasCo (expert-curated database, (De Boer and Hughes 2012)), using a pseudocount of 0.001. Binding site scores are the log-likelihood of observing the sequence under the motif model compared to a background model of nucleotide frequencies (G+C = 34.2% for *S. cerevisiae* and 36.3% for *S. uvarum*). For each gene we used the highest scoring binding site within its upstream intergenic region. Negative scores were set to zero. The temperature effects of *S. cerevisiae* alleles were used in the following analysis. Binding sites associated with temperature effects were identified by linear regression with the average binding site score of the two species. Mann-Whitney tests were used to assess enrichment of binding sites in temperature-responsive genes compared to genes without a

temperature response. Motif models that were significant for both linear regression and Mann-Whitney tests after Holm-Bonferroni correction were considered positive hits.

Predicted nucleosome occupancy was generated by NuPoP (Xi et al. 2010), using the yeast model for both species. The average nucleosome occupancy across each promoter was used. For each intergenic region, we calculated a weighted score: the average binding site score of the two species \* (1- nucleosome occupancy of *S. cerevisiae* promoter). Linear regression and Mann-Whitney tests were used to predict temperature effects by the weighted scores.

Binding site divergence for each binding site model was calculated by the difference between the highest scoring site for each allele. To test for associations between expression and the combined divergence of all binding sites we used the average of the absolute value of binding site divergence. For each motif model, linear regression was used to test association between binding site divergence and allele specific effects.

## **Results**

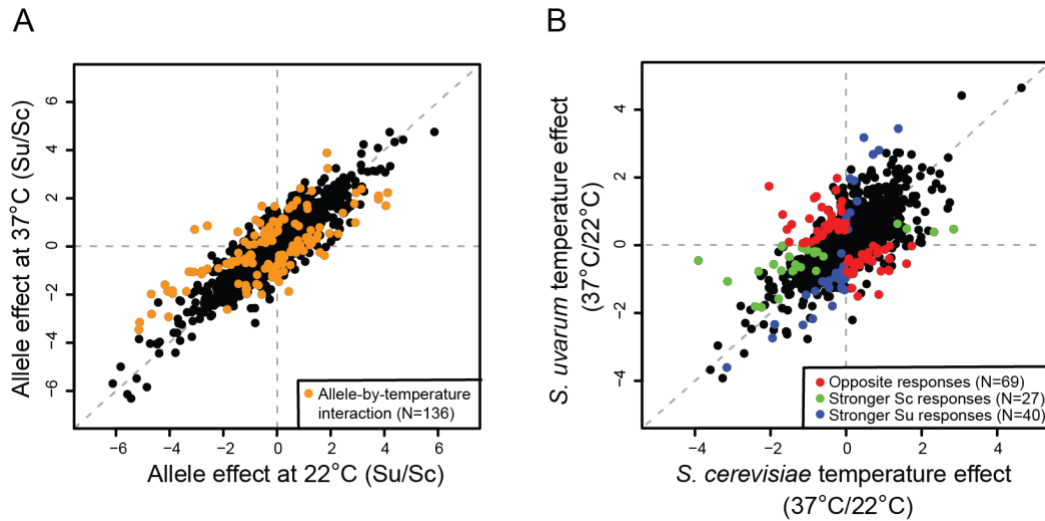
### *Effects of temperature on allele-specific expression*

To measure the effects of temperature on allele-specific expression (ASE) we generated RNA-seq data from an *S. cerevisiae* × *S. uvarum* hybrid during log phase growth at low (22°C), intermediate (33°C) and high (37°C) temperatures. Out of 5,055 orthologs, we found 2,950 (58%) that exhibited allele-specific expression, 1,669 (33%) that exhibited temperature-dependent expression and 136 (2.7%) that exhibited allele-by-temperature interactions (FDR < 0.05). For the 1,669 temperature-responsive genes, expression levels were highly correlated between 33°C and 37°C (Pearson's correlation coefficients = 0.97) and 37°C was more different



from 22°C than 33°C (Pearson's correlation coefficients = 0.89 for 22°C-37°C, 0.93 for 22°C-33°C). Despite the abundant temperature responses, allele differences were similar across temperatures with Pearson's correlation coefficients of 0.90, 0.92 and 0.96 for 22-37°C, 22-33°C and 33-37°C, respectively (Fig. 2-2A). In addition, the proportion of genes with the *S. cerevisiae* allele expressed at higher levels than the *S. uvarum* allele was 49.9, 50.7, 49.8% at 22, 33 and 37°C, respectively. Thus, there is no tendency toward higher *S. cerevisiae* allele expression at high temperature. *S. uvarum* alleles can be expressed at the same level as their *S. cerevisiae* ortholog at 37°C despite the fact that these promoters don't experience high temperature in *S. uvarum* due to its temperature restriction.

Allele-specific temperature responses may reflect *cis*-regulatory changes involved in thermal differentiation. We therefore examined the 136 genes with a significant temperature-by-allele interaction. The gene set is not enriched for any GO terms ( $p > 0.05$ ) and contains genes involved in a variety of biological processes. However, four genes are involved in trehalose metabolic process (*NTH2*, *TPS2*, *HSP104*, *PGM2*), and trehalose has been shown to influence thermotolerance (Eleutherio et al. 1993). Among the 136 genes, we found 27 where the *S. cerevisiae* allele responded to high temperature (37°C) more strongly than *S. uvarum*, and 40 genes where the *S. uvarum* allele responded more strongly. In the remaining 69 genes, alleles from the two species showed responses in opposite directions (Fig. 2-2B).

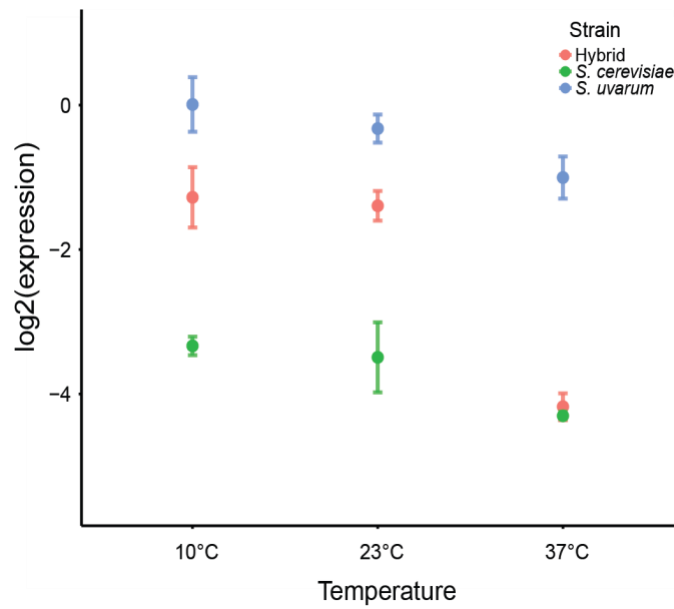


**Figure 2-2. Temperature-dependent allele effects.** The 136 genes with temperature-dependent allele effects are shown in color (legend) compared to all other genes (black, N = 4,919). **A.** Species' allele effects (*S. uvarum*/*S. cerevisiae*) at low versus high temperature. **B.** Temperature effects (37°C/22°C) of *S. cerevisiae* (Sc) versus *S. uvarum* (Su). Temperature effects are classified into those with species' alleles have an opposite temperature response (red), the *S. cerevisiae* allele responding to temperature more strongly than *S. uvarum* (green), and the *S. uvarum* allele responding to temperature more strongly than *S. cerevisiae* (blue).

#### *Effects of temperature on hybrid gene expression*

To characterize temperature-dependent changes in gene expression we examined 211 genes that showed both a significant temperature effect (FDR < 0.05) and a 2-fold or more difference between the low (22°C) and high (37°C) temperatures. Unexpectedly, genes expressed at higher levels at the low temperature were enriched for genes involved in protein folding (*AHA1*, *MDJ1*, *BTN2*, *SSA2*, *HSP104*, *HSC82*, *SIS1*, *STI1*, *HSP82*, *CUR1*,  $p = 0.00829$ , Table S2-1). Typically, protein chaperones are induced in response to heat stress or misfolded proteins (Verghese et al. 2012).

To confirm the higher expression of genes involved in protein folding at 22°C and test whether this expression is specific to the hybrid or also found in one of the parents, we examined *HSP104* expression by quantitative PCR (Fig. 2-3). Similar to our RNA-seq data, in the hybrid *HSP104* is expressed at higher levels at low temperatures (10°C and 23°C) compared to high temperatures (37°C) (5-fold change,  $p = 0.0006$ , t-test). Consistent with prior work (Gasch et al. 2000), in both parental species *HSP104* is expressed at the same level across temperatures and any transient induction that might have occurred upon a shift to 37°C is no longer present (linear regression,  $p = 0.11$  for *S. cerevisiae* and 0.13 for *S. uvarum*). However, in *S. uvarum* *HSP104* is expressed at higher levels than *S. cerevisiae* across all temperatures (t-test,  $p = 0.007$ , 0.013, 0.006 for 10°C, 23°C and 37°C, respectively). The atypical pattern of *HSP104* expression in the hybrid can be explained by a change in the dominant trans-acting environment. At low temperatures (10°C and 23°C) *S. uvarum* tends to dominate the trans-environment leading to high levels of *HSP104* expression whereas at 37°C *S. cerevisiae* completely dominates the trans-environment leading to low levels of *HSP104* expression.



**Figure 2-3. Temperature dependent *HSP104* expression in *S. cerevisiae*, *S. uvarum* and their hybrid.** Expression is based on qPCR with points showing the mean and bars the standard errors. Hybrid expression is the sum of the two alleles.

#### *Test for temperature-specific directional evolution*

Under a neutral model with no change in the selective constraints on gene expression, allele-specific differences in gene expression between species are expected to be symmetrically distributed. Parallel directional changes in gene expression among a group of functionally related or co-regulated genes can reflect selection (Bullard et al. 2010; Fraser 2011). Such groups have been reported in a hybrid of *S. cerevisiae* and *S. uvarum* by a sign test (Bullard et al. 2010), but the phenotypic consequences of these expression changes are not known. We tested whether patterns of directional selection are temperature-dependent, as might be expected if they are related to thermal differentiation. For example, consistent higher expression of the *S. cerevisiae* allele at the high but not low temperature would implicate directional selection in temperature-

dependent expression divergence. We applied the sign test to ASE at each temperature separately and found 8, 9 and 13 groups of genes with directional ASE at 22°C, 33°C and 37°C, respectively ( $p < 0.01$ , FDR = 0.27, 0.24, 0.068 for 22°C, 33°C, 37°C, respectively; Table 2-1). Seven groups are significant for all three temperatures, including the previously reported histidine biosynthesis and lysine biosynthesis groups (Bullard et al. 2010). Although we found a few groups specific to one or two temperatures using the  $p < 0.01$  cutoff (e.g. Cluster\_MET31, Cluster\_adata-CalciumSpecific, etc.), all of these groups showed similar sum of scores across temperatures and  $p < 0.10$  (Table 2-1). Therefore, none of the groups exhibiting directional divergence are temperature-specific.

Table 2-1. Groups of genes showing directional evolution at three temperatures.

Group <sup>1</sup>	Number of genes in the group <sup>2</sup>	Sum of scores <sup>3</sup>			Annotation <sup>4</sup>
		22°C	33°C	37°C	
Cluster_FHL1	93 (89)	29**	27**	17*	Ribosomal proteins
Cluster_RPs	136 (129)	38**	40**	29**	Ribosomal proteins
Cluster_Histidine	8 (6)	6**	5*	4*	Histidine biosynthesis
Node 39	36 (24)	-12**	-10*	-8*	Organonitrogen catabolism
Cluster_MET31	17 (17)	10*	6	6*	Amino acid metabolism
Cluster_Lysine	9 (9)	6*	7**	5*	Lysine biosynthesis
Cluster_TFs	18 (12)	-7*	-7*	-5*	Transcription factors
Node 81	7 (7)	5*	5*	5**	Lysine biosynthesis
Cluster_adata-CalciumSpecific	71 (44)	-10	-15**	-15**	Membrane localization
Node 36	182 (144)	-20	-17	-19*	Unknown

Cluster_SWI6	29 (28)	-7	-6	-8*	Cell cycle regulation
Node 8	83 (64)	-10	-7	-11*	Oxidative stress
Cluster_PHO4	14 (12)	5	5	5*	Unknown
Node 87	8 (5)	4	4*	3	Microtubule polymerization

<sup>1</sup> Groups are defined by (Gasch et al. 2004).

<sup>2</sup> Number of genes with available data is shown in parentheses.

<sup>3</sup> Positive scores indicate *S. cerevisiae* alleles are expressed higher than *S. uvarum* alleles; negative scores indicate *S. cerevisiae* alleles are expressed lower than *S. uvarum* alleles. Significant groups are indicated for  $p < 0.01$  (\*) and  $p < 0.001$  (\*\*).

<sup>4</sup> The groups are annotated based on GO terms of genes in the group.

#### *Promoter changes associated with expression divergence*

To identify promoter features that could explain allele-specific differences in expression we examined intergenic substitution rate, transcription factor binding site scores and their interaction with nucleosome occupancy. Among ASE genes, intergenic substitution rates were weakly correlated with gene expression divergence (Spearman's  $\rho = 0.064$ ,  $p = 0.002$ ). Given these differences we also calculated rates of binding site divergence using binding sites scores from 244 transcription factor binding site models (De Boer and Hughes 2012) and found a weak correlation between expression divergence and binding site divergence (Spearman's  $\rho = 0.05$ ,  $p = 0.0119$ ).

To identify binding sites that could explain allele-specific expression we first tested each binding site model for its ability to predict temperature responsive genes (22°C vs 37°C). We identified 17 motifs associated with genes induced at 22°C and 13 motifs associated with genes

repressed at 22°C (Holm–Bonferroni corrected  $p < 0.05$  for both linear regression and Mann-Whitney test, Fig. S2-1). Many of the motifs (11/17) associated with up-regulated genes are similar to the stress response element (AGGGG), including the canonical stress response factors *MSN2* and *MSN4*. Other motifs known to be involved in the stress response include the heat shock factor *HSF1*, which is consistent with the observed up-regulation of heat shock genes at 22°C (Table S2-1). Motifs enriched upstream of down-regulated genes are involved in glucose repression, e.g. *MIG1*, *MIG2*, *MIG3* and *ADR1*. *UME6* was also found, consistent with down-regulation of meiotic genes at 22°C revealed by GO analysis (Table S2-1). We also examined the correlations using a weighted score that accounts for both TF binding and nucleosome occupancy (see Methods), but the correlations were not greatly improved with the nucleosome weighted binding site scores.

Given the motifs associated with the temperature response, we tested each motif for an association between binding site divergence and ASE at 22°C. Within genes down-regulated at 22°C, divergence of 5 motifs was found to have a weak but significant association with expression divergence (*MIG1*, *MIG2*, *MIG3*, *TDA9* and *YGR067C*, linear regression, Holm-Bonferroni corrected  $p < 0.05$ , Fig. S2-1). No motifs were correlated with ASE in genes up-regulated at 22°C, but one motif (*ARO80*) was correlated with ASE in genes that showed allele-by-temperature effects and up-regulation at 22°C ( $p = 0.0028$ , adjusted r-squared = 0.11). The weak correlations suggest that ASE is likely often caused by *cis*-regulatory mutations outside of known binding sites.

## Discussion

Environment-dependent gene expression is likely an important component of fitness. While *cis*-acting divergence on gene expression is abundant between species, the extent to which these *cis*-effects are environment-dependent is not often known. In this study, we show most *cis*-effects are independent of temperature in two thermally diverged yeast species. Further, we find that most *S. uvarum* alleles are expressed at levels similar to *S. cerevisiae* alleles at 37°C, even though *S. uvarum* does not grow at this temperature. Below, we discuss these results in relation to prior studies of variation in gene expression across environments and discuss the challenge of identifying changes in promoter sequences responsible for divergence in gene expression.

### *Environment-dependent cis-effects*

Changes in gene regulation may be an important aspect of how species adapt to different environments. Although there is extensive variation in gene expression-by-environment interactions (Hodgins-Davis and Townsend 2009), the extent to which these differences are caused by *cis*- or *trans*-acting factors is not as well characterized. We find that most *cis*-effects do not depend on temperature, only 136 of the 2,950 genes exhibiting ASE show temperature-dependent ASE. Thus, even though *S. uvarum* promoters have never been exposed to high temperatures, they can drive expression levels similar to those of *S. cerevisiae*. The consistent *cis*-effects across temperatures suggest that most *cis*-regulatory divergence is not associated with thermal divergence between the two species. Previous studies also found that *cis*-effects tend to be constant across environments and only a small subset of them are environment-dependent (Smith and Kruglyak 2008; Tirosh et al. 2009; Fear et al. 2016; He et al. 2016). Although we did not examine *trans*-effects genome-wide, the shift in the *trans*-effect of *HSP104* with temperature



is consistent with prior work showing that *trans*-effects play a more pronounced role in environment-dependent differences in gene expression (Smith and Kruglyak 2008; Tirosh et al. 2009).

Although only a small number of genes showed a significant allele-by-temperature interaction, some may be relevant to thermal differentiation. Of particular interest are genes where the *S. cerevisiae* but not the *S. uvarum* allele responded to temperature. One noteworthy example of such is *TPS2*, which showed 2.5- compared to 1.5-fold induction of the *S. cerevisiae* compared to the *S. uvarum* allele, respectively. *TPS2* is involved in trehalose biosynthesis and essential to heat tolerance in *S. cerevisiae* (De Virgilio et al. 1993). The lower *cis*-regulatory activity of *TPS2* in *S. uvarum* might cause lower levels of trehalose and compromise its heat tolerance. In addition, three other genes (*NTH2*, *HSP104*, *PGM2*) in the trehalose pathway also showed allele-by-temperature effects, suggesting that the transcriptional regulation of this pathway might have diverged in the two species.

Among the 136 genes with temperature-dependent ASE, 67 genes showed a consistent direction but different magnitude of response for the two species' alleles. The majority of them (53) were differentially induced at 22°C compared to 37°C and many are known to be induced by heat (*PIC2*, *SSE2*, *YKL151C*, *SIS1*, *IKS1*, *AHA1*, *EDC2*, *GSY2*, *HSP104*, *PUN1*, *TPS2*), oxidative stress (*ZWF1*, *YPR1*, *SOD1*) or other stresses (*CMK2*), consistent with the hybrid exhibiting a stress response at 22°C. However, there is no bias for the *S. cerevisiae* or the *S. uvarum* allele being more induced (23 vs. 30 genes). In addition, in several heat-related genes (*AHA1*, *GSY2*, *HSP104*), the *S. cerevisiae* allele is more induced at 33°C than the *S. uvarum* allele, but at 22°C they are equally induced (with expression levels higher than or equal to those

at 33°C). However, interpreting these changes is difficult given the *trans*-acting stress response is strongest at 22°C.

The 69 genes with alleles showing opposite responses to temperature are also worth discussing as some of them might be indicative of misregulation or thermal divergence. We examined their ASE pattern at 22°C and 37°C and classified them based on: ASE at both temperatures (24 genes), ASE at one temperature (42 genes), or ASE at neither of the two temperatures (3 genes). Among the 66 genes that showed ASE at one or more temperatures, only two genes (*IMP2*, *POR2*) showed ASE at both temperatures but with opposite allele effects, where the *S. cerevisiae* alleles were higher than the *S. uvarum* alleles at 22°C but lower at 37°C. For the rest of the 64 genes, they either had ASE at both temperature but one allele consistently higher than the other allele, or showed ASE at one temperature but not the other. Thus, the 64 genes can be classified into two groups with 22°C-divergent or 37°C-divergent expression patterns. Two-thirds of them (43 genes) showed larger allele differences at 22°C than 37°C, i.e. 22°C-divergent. Among these genes, the *S. cerevisiae* alleles were expressed higher than the *S. uvarum* alleles at 22°C in 22 genes, vice versa for the remaining 21 genes. Interestingly, many genes in this group are related to mitochondrial function or oxidative stress (*GAD1*, *TIR3*, *QRI7*, *AIM41*, *YIG1*, *LAM4*, *YKL162C*, *THI73*, *ARG7*, *ICY1*, *YJL193W*, *YNL200C*, *YNL144C*, *YNL208W*). Mitochondrial function has been shown to be related to *S. cerevisiae*'s thermotolerance (Davidson and Schiestl 2001); thus the *cis*-acting divergence in mitochondria-related genes might be important to thermal divergence. In addition, the hybrid strain used in this study carries only *S. cerevisiae* mitochondrial DNA (mtDNA). Although it is also possible that the responses of mitochondria-localized genes are affected by *S. cerevisiae* mtDNA, this would imply species-specific feedback regulation on mRNA levels.

Besides the mitochondrial genes, membrane proteins (*YLR046C*, *YJR015W*, *THI73*), cell wall (*TIR3*, *CWPI*) and mating-related genes (*PRM4*, *AXL1*, *SIR1*) were also found in the 22°C-divergent group. The 21 genes in the 37°C-divergent group are involved in responses to glucose limitation (*GTT1*, *GSY1*), sporulation (*QDR3*, *NPPI*), cell signaling (*RHO5*, *TOS3*, *ROM1*), nutrient metabolism (*QDR3*, *YJR124C*, *NPPI*, *STR2*, *ATF2*) and mitochondrial functions (*TOS3*).

Taken together, one of the most notable features of the allele responses is that they more often diverge at 22°C than 37°C (43 vs. 21). Given that expression at 22°C resembles a stress response (Table S2-1), the greater divergence at 22°C may reflect divergent stress responses between the two species. Although the genes with allele-specific temperature responses have diverse biological functions, the stress- and mitochondrial-related genes are more often differentially induced at 22°C. However, it is also important to consider that these differences may only be present in a hybrid environment where we find a stronger stress response at low compared to high temperature.

#### *Unexpected heat shock response at low temperatures*

The non-canonical expression of heat shock genes at 22°C is somewhat perplexing. Because we measured expression at constant temperatures we did not expect to see induction of heat shock genes, which normally occurs within 30 minutes of treatment and then dissipates (Gasch et al. 2000). Given the high expression level of *HSP104* in *S. uvarum* across all temperatures, one potential explanation for the heat shock response is a *trans*-signal produced by the *S. uvarum* genome. The absence of the heat shock response in the hybrid at high temperature may be a consequence of loss of the *S. uvarum trans*-signal, although this does not explain the

high *HSP104* expression at high temperature in *S. uvarum*. Sample mix-up is unlikely as the *HSP104* experiment was done independently and is consistent with the original RNA-seq experiment.

The heat shock gene expression profile shows that the hybrid is under stress at 22°C but not 37°C. To better understand this counterintuitive phenomenon, we compared the hybrid expression profile to previously published *S. cerevisiae* (Gasch et al. 2000) and *S. uvarum* (Caudy et al. 2013) datasets. The hybrid temperature effect (37°C over 22°C) associates with 285 of 477 stress responses of either *S. cerevisiae* or *S. uvarum* (Spearman's correlation test, Holm-Bonferroni corrected  $p < 0.05$ ). However, 232 of the 285 correlations are negative, implying that 22°C is more stressful than 37°C in the hybrid. Interestingly, the strongest positive correlation is between the hybrid's temperature response and *S. uvarum*'s 17°C to 30°C response at 60 min (Spearman's  $\rho = 0.23$ , Holm-Bonferroni corrected  $p = 5.39E-48$ ). In contrast, the correlations with *S. uvarum*'s 25°C to 37°C or 25°C to 42°C response are negative. Similar to the hybrid, heat shock genes are expressed higher at 17°C than 30°C in *S. uvarum*, but the pattern is not seen in the other two temperature shifts (Caudy et al. 2013). These differential correlations indicate *S. uvarum*'s heat shock response may be sensitive to specific temperatures used in the shifts. However, it is also important to note that heat shock proteins are not specific to temperature changes but are part of the general environmental stress response which can be induced by any number of environmental changes (Gasch et al. 2000). Taken together, the stress response induced in the hybrid at 22°C may reflect a contribution from the non-canonical temperature response in *S. uvarum*.

## *Signatures of selection on cis-acting divergence in gene expression*

The sign test of allele imbalance across functionally related genes has been used in a variety of configurations to detect polygenic *cis*-regulatory adaptation (Bullard et al. 2010; Fraser et al. 2010; Fraser 2011; Naranjo et al. 2015; He et al. 2016). However, previous applications of the test were to expression levels under standard growth conditions. Because gene expression is environment-dependent, some signals of selection may only be uncovered by examining expression in environments to which an organism adapted. However, our results indicate that directional ASE as found by the sign test is not temperature-dependent, consistent with our observation that most *cis*-effects are not temperature dependent. Our results do not rule out the possibility of *trans*-acting expression differences important to thermal differentiation, nor do they address *cis*-acting changes that occur immediate after a temperature shift and which are typically much stronger and more widespread than those that persist for hours after the initial shift (Gasch et al. 2000).

In addition to the histidine and lysine biosynthesis groups reported by Bullard et al. (2010), we found several other groups of genes showing a signature of directional evolution. Among these, the ribosomal genes show a strong bias toward higher *S. cerevisiae* allele expression (Table 2-1), which could indicate a difference in translational capacity of the two species. Two other groups, Node 39 (organonitrogen catabolism) and Cluster\_MET31 (amino acid metabolism) provide new evidence for divergence in nutrient metabolism between the two species.

Most groups identified by the sign test contain a substantial number of temperature-responsive genes, with the lysine biosynthesis pathway showing the highest fraction (8/9). The

pathway consists of nine genes (*LYS1*, *LYS2*, *LYS4*, *LYS5*, *LYS9*, *LYS12*, *LYS14*, *LYS20*, *LYS21*), eight of which are induced at 22°C, with *LYS4* and *LYS20* showing allele-by-temperature effects. The *S. cerevisiae* allele of *LYS20* is induced at 22°C more than the *S. uvarum* allele (3.2 vs. 1.0 fold). Although not a significant temperature-by-allele interaction, a similar pattern is present for *LYS1* (4.1 vs. 2.9 fold) and *LYS2* (2.3 vs. 2.1 fold). The weak responses of *S. uvarum* alleles might reflect deficiency in activating the lysine biosynthesis pathway at a given temperature or under stress, which is critical for amino acid homeostasis. Also, the lysine biosynthesis pathway is known to be induced by mitochondrial retrograde signaling in response to compromised mitochondrial respiratory function (Liu and Butow 2006) and could potentially be affected by the *S. cerevisiae* mtDNA.

#### *Binding sites are only weakly related to expression divergence*

Consistent with previous reports (Tirosh et al. 2008; Tirosh and Barkai 2008; Chen et al. 2010; Zeevi et al. 2014), we only found weak correlations between binding site changes and allele-specific expression. Previous work has shown that binding sites in nucleosome depleted regions are more likely to cause changes in gene expression (Swamy et al. 2011). Yet, incorporation of predicted nucleosome occupancy did not improve our ability to predict gene expression. This finding is consistent with another study that found no relationship between divergence in nucleosome occupancy and gene expression in yeast (Tirosh et al. 2010). One explanation for the weak correlations is that ASE may often be caused by *cis*-regulatory mutations outside major binding sites, e.g. (Levo et al. 2015). Genes in the lysine biosynthesis pathway provide a good example of conserved binding sites: seven genes in the pathway showed higher *S. cerevisiae* expression, yet binding sites for *LYS14*, the major transcription factor that

regulates these genes (Becker et al. 1998), are conserved in all of them. Furthermore, the lysine genes are also not enriched for divergence in other motifs present upstream of these genes (e.g. *MOT2*, *XBP2*, *RTG1*, *RTG3*,  $p > 0.05$ , Mann-Whitney test).

Despite binding site divergence being only weakly related to ASE, we found a few significant associations with specific binding sites. One of these, *ARO80* sites, correlated with temperature-dependent expression differences largely due to two genes *ARO9* and *ARO10* (Fig. S2-1, S2-2). In both cases, the *S. uvarum* promoters have lower binding scores and lower expression of the *S. uvarum* allele (Fig. S2-2). Interestingly, the number of monomers in the *ARO80* binding sites also differs between *S. cerevisiae* and *S. uvarum*. In both genes, *S. cerevisiae* sites are tetrameric and *S. uvarum* sites are trimeric (Fig. S2-2). The example of *ARO80* suggests expression divergence might associate with changes in the number of binding sites, which our binding site analysis didn't consider.

# **Chapter 3: Mitochondria-encoded genes contribute to evolution of heat and cold tolerance in yeast**

This chapter was done in collaboration with David Peris, Chris Todd Hittinger, Elaine A. Sia and Justin C. Fay, and is a reprint from a manuscript accepted by *Science Advances*.



## Abstract

Genetic analysis of phenotypic differences between species is typically limited to interfertile species. Here, we conduct a genome-wide non-complementation screen to identify genes that contribute to a major difference in thermal growth profile between two reproductively isolated yeast species, *Saccharomyces cerevisiae* and *S. uvarum*. The screen revealed a single nuclear-encoded gene, but a large effect of mitochondrial DNA (mitotype) on both heat and cold tolerance. Recombinant mitotypes indicate multiple genes contribute to thermal divergence and we show that protein divergence in *COXI* affects both heat and cold tolerance. Our results point to the yeast mitochondrial genome as an evolutionary hotspot for thermal divergence.

## Introduction

The genetic architecture of phenotypic divergence between species is unresolved. There remains considerable uncertainty as to whether evolution occurred through accumulation of numerous small-effect changes (“micromutationism”) or often involves “major genes” of large effect (Orr and Coyne 1992). While quantitative trait mapping has been successfully applied to closely related, interfertile species (reviewed in (Orr 2001)), the results may not be representative of phenotypic divergence in general, since the characters that distinguish sibling species and domesticated organisms evolved over short time-scales and potentially favor large-effect loci. However, systematic dissection of divergence between distantly related species has been difficult due to reproductive barriers.

The genus *Saccharomyces* contains post-zygotically isolated species with substantially diverged genomes, and the ease of genetic manipulation of yeast may allow us to address the genetic architecture of evolution with a systematic approach. While the *Saccharomyces* species share their preference for fermentative metabolism with many other yeast species (Hagman et al. 2013), they differ dramatically in their thermal growth profile (Gonçalves et al. 2011; Salvadó et al. 2011). *S. cerevisiae* is the most heat-tolerant species in this lineage, capable of growing at temperatures of 37-42°C, while its sister species *S. paradoxus* can grow up to 39°C and the more distantly related *S. kudriavzevii* and *S. uvarum* are more cold-tolerant and only capable of growing at temperatures up to 34-35°C (Gonçalves et al. 2011; Salvadó et al. 2011). Previous studies in yeasts have implicated a small number of genes involved in temperature divergence (Gonçalves et al. 2011; Paget et al. 2014). However, every gene product has the potential to be thermolabile, and only a single systematic screen has been conducted (Weiss et al. 2018), which

reported that multiple genes contribute to thermal differences between *S. cerevisiae* and *S. paradoxus*, two species with modest differences in heat tolerance.

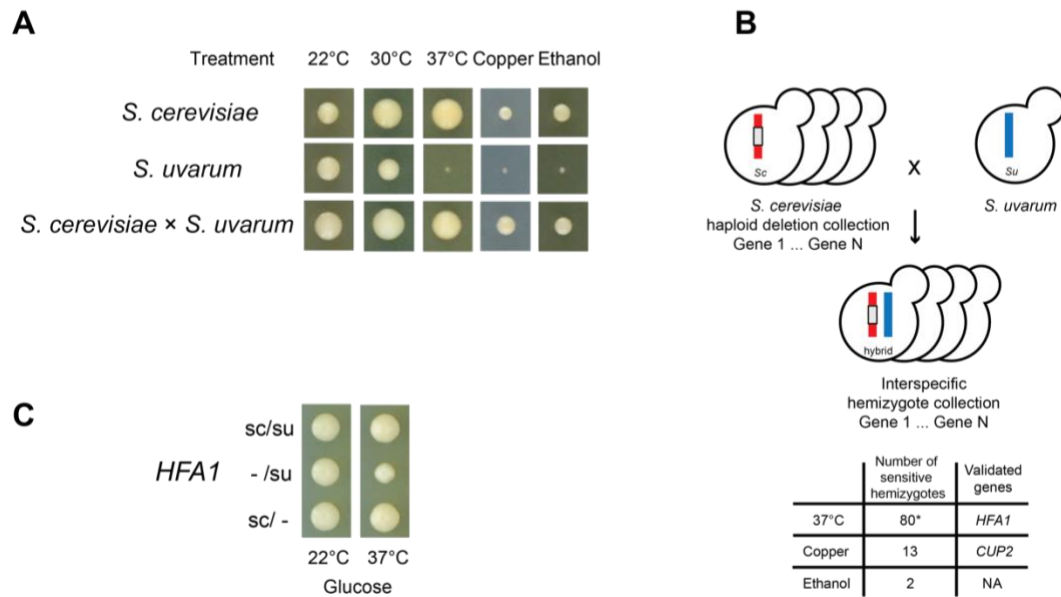
In the present study, we examined the genetic basis of thermal divergence between *S. cerevisiae* and *S. uvarum*, two species that are more different at synonymous sites than human and mouse (Waterston et al. 2002; Kawahara and Imanishi 2007). These two species are capable of forming hybrids, but the hybrids cannot produce viable spores. Mechanisms underlying the reproductive isolation could involve mitochondrial-nuclear incompatibilities (Lee et al. 2008; Chou et al. 2010), defects in recombination due to high levels of sequence divergence (Hunter et al. 1996; Liti et al. 2006), and chromosomal rearrangements (Fischer et al. 2001; Delneri et al. 2003). Of relevance, mitochondrial genome variation has been shown to impact high temperature growth in *S. cerevisiae* (Paliwal et al. 2014; Wolters et al. 2018) and *S. paradoxus* (Leducq et al. 2017).

To identify genes involved in the evolution of thermal growth differences, we screened 4,792 non-essential genes for non-complementation and used the reciprocal hemizyosity test (Steinmetz et al. 2002) to validate genes that came out of the screen. While no single genes of large effect were recovered, we found that mitochondrial DNA (mtDNA) plays a remarkable role in divergence of both heat and cold tolerance across the *Saccharomyces* species and that multiple mitochondria-encoded genes are involved, including *COXI*, previously shown to be involved in mitochondrial-nuclear interspecific incompatibilities (Chou et al. 2010).

## Results

### *A non-complementation screen for thermosensitive alleles reveals mitochondrial effects*

Hybrids of *S. cerevisiae* and *S. uvarum* are heat tolerant (Fig. 3-1A). Thus, deletion of *S. cerevisiae* heat tolerant alleles in a hybrid should weaken heat tolerance through non-complementation. We screened 4,792 non-essential genes in the yeast deletion collection for such thermotolerance genes by mating both the *MATa* (BY4741) and *MAT $\alpha$*  (BY4742) deletion collection to *S. uvarum* and growing them at high temperature (37°C). For comparison, we also screened the resulting hemizygote collections for two other traits where the *S. cerevisiae* phenotype is dominant in the hybrid (Fig. 3-1A): copper resistance (0.5 mM copper sulfate) and ethanol resistance (10% ethanol at 30°C). We found 80, 13, and 2 hemizygotes that exhibited reduced resistance to heat, copper, and ethanol, respectively, in both the BY4741 and BY4742 hemizygote collections (Fig. 3-1B). In our initial assessment of these genes, we validated a copper-binding transcription factor, *CUP2* (Buchman et al. 1989), for copper resistance through reciprocal hemizyosity analysis (Fig. S3-1).



**Figure 3-1. A non-complementation screen identified genes underlying phenotypic divergence between *S. cerevisiae* and *S. uvarum*.** (A) *S. cerevisiae* and *S. uvarum* differ in heat (37°C), copper (0.5mM, 22°C), and ethanol (10%, 30°C) tolerance. The resistant *S. cerevisiae* alleles are dominant, shown by the hybrid (*S. cerevisiae* × *S. uvarum*) compared to *S. cerevisiae* (diploid, S288C background) and *S. uvarum* (diploid, CBS7001 background). Growth is after 3 days. (B) *S. cerevisiae* haploid deletion collection was crossed to *S. uvarum* to construct an interspecies hemizygote collection. The number of non-complementing genes is shown for each phenotype; the asterisk indicates that the number includes strains carrying *S. uvarum* mtDNA. (C) *HFA1* hemizygote with only an *S. cerevisiae* allele (sc/-) shows better 37°C growth than one with only an *S. uvarum* allele (-/su). Growth is after 5 days. See Fig. S3-1B for quantification.

Nearly all of the heat-sensitive hemizygotes (77/80) were from respiration-deficient (“petite”) *S. cerevisiae* parents. We found many of these strains carried *S. uvarum* mitochondrial DNA (mtDNA) via PCR of a mitochondrial marker. Although not extensively tested, other hemizygotes are expected to carry *S. cerevisiae* mtDNA, a typical outcome of *S. cerevisiae* × *S. uvarum* crosses (Albertin et al. 2013). The difference in mtDNA inheritance was likely caused by loss of mtDNA in the *S. cerevisiae* petite parents. We confirmed one gene (*HFA1*) by reciprocal hemizyosity analysis (Fig. 3-1C, Fig. S3-1) that causes a moderate loss of heat tolerance due to the *S. uvarum* allele in the presence of *S. cerevisiae* mtDNA. *HFA1* encodes a mitochondrial acetyl-coenzyme A carboxylase and is involved in mitochondrial fatty acid biosynthesis (Hoja et al. 2004).

The inheritance of *S. uvarum* mtDNA in heat-sensitive hemizygotes suggested that mtDNA, rather than the deletion, could be the cause. To test whether the species' mtDNA (“mitotype”) affects heat tolerance, we generated diploid hybrids of wild-type *S. cerevisiae* and *S. uvarum* with reciprocal mitotypes and grew them at different temperatures. In comparison to the hybrid with *S. cerevisiae* mitotype, the hybrid with the *S. uvarum* mitotype showed reduced fermentative growth (glucose medium) at 37°C compared to 22°C and almost no respiratory growth (glycerol medium) at 37°C (Fig. 3-2A).

*S. uvarum* is not only known to be heat sensitive, but also exhibits enhanced growth at low temperatures relative to *S. cerevisiae* (Gonçalves et al. 2011). We thus tested and found that *S. uvarum* mitotype conferred a growth advantage at 4°C in comparison to *S. cerevisiae* mitotype (Fig. 3-2A), suggesting a potential trade-off between the evolution of heat and cold tolerance.

To test whether mtDNA-mediated evolution of temperature tolerance is specific to either the *S. cerevisiae* or *S. uvarum* lineages, we generated five additional hybrids with both parental mitotypes using two other *Saccharomyces* species (Fig. S3-2). In comparison to the 22°C control, we find that both the *S. cerevisiae* and *S. paradoxus* nuclear genome conferred heat tolerance to hybrids with *S. kudriavzevii* and *S. uvarum* (rho<sup>o</sup> comparison), but the *S. cerevisiae* mitotype conferred heat tolerance in comparison to the *S. paradoxus*, *S. kudriavzevii*, and *S. uvarum* mitotypes on glucose medium. For cold tolerance we find that the *S. uvarum* mitotype conferred greater cold tolerance relative to the *S. cerevisiae*, *S. paradoxus*, and *S. kudriavzevii* mitotypes. Interestingly, none of the hybrids was as cold tolerant as *S. uvarum* on glycerol. Our results suggest that mtDNA has played an important role in divergence of thermal growth profiles among the *Saccharomyces* species, with heat tolerance evolving primarily on the lineage leading to *S. cerevisiae* and cold tolerance evolving primarily on the lineage leading to *S. uvarum*. A related study has shown these differences have had a direct impact on the domestication of lager-brewing yeast hybrids to low-temperature fermentation (Baker et al. 2018).

#### *Recombinant analysis identifies contribution of multiple mitochondria-encoded genes*

To identify mtDNA genes conferring heat tolerance to *S. cerevisiae*, we tested whether *S. uvarum* alleles can rescue the respiratory deficiency of *S. cerevisiae* mitochondrial gene knockouts at high temperature. We crossed *S. uvarum* to previously constructed *S. cerevisiae* mitochondrial knockout strains and plated them on glycerol medium at 37°C. Because heteroplasmy is unstable in yeast, this strategy selects for recombinants between the two mitochondrial genomes: *S. uvarum* mtDNA is needed to rescue the *S. cerevisiae* deficiency, and

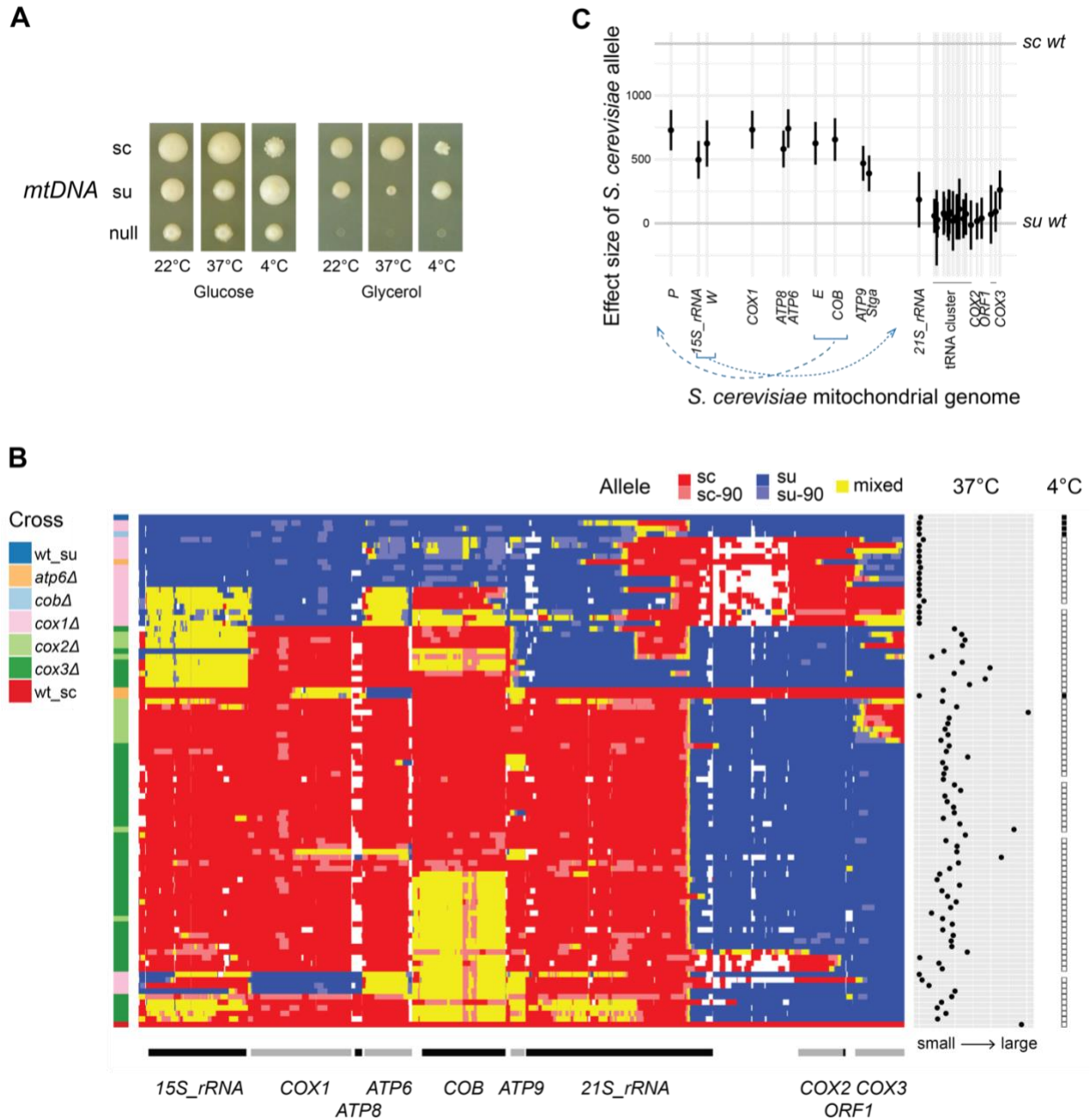
*S. cerevisiae* mtDNA is needed to grow at high temperature (Fig. S3-3). If the *S. uvarum* gene required for *S. cerevisiae* rescue is temperature sensitive, we expect to see no or small colonies on 37°C glycerol plates. Of the six genes tested, *COX2* and *COX3* deletions were rescued by *S. uvarum* at high temperature, although the colonies were often smaller than the hybrid with wild-type *S. cerevisiae* mtDNA. In contrast, *COX1* and *ATP6* deletions were minimally rescued (Fig. 3-2B), and *COB* and *ATP8* deletions were not rescued. However, the absence of rescue could also result from a lack of recombination, especially for *COB* because its genomic location has moved between the two species.

Using genome sequencing, we mapped breakpoints in 90 recombinants to determine which *S. cerevisiae* genes are associated with high temperature growth. The recombinants showed hotspots at gene boundaries and within the 21S ribosomal RNA (Fig. 3-2B). In most cases, the two species' mtDNA recombine into a circular mitochondrial genome, but sometimes recombination resulted in mitochondrial aneuploidy, particularly for regions where the two species' mitochondrial genomes are not co-linear (see Fig. S3-4B for examples). One complication of measuring mtDNA-dependent heat tolerance is the high rate of mtDNA loss, typically 1% in *S. cerevisiae* strains, but much higher in the hybrids and variable among recombinants (Supplementary text, Fig. S3-5). We thus measured the frequency of petites at 22°C and heat tolerance by the size of single colonies at 37°C on glycerol. We found that the petite frequency was associated with the absence of *S. cerevisiae* *ORF1* (F-*SceIII*, (Peris, Arias, et al. 2017), a homing endonuclease linked to *COX2* (Fig. S3-5B & C). For heat tolerance, we found a region including four protein-coding genes (*COX1*, *ATP8*, *ATP6*, and *COB*) with the largest effect (Fig. 3-2C). The effects associated with these genes are small compared to the total difference between two wild-type mitotypes, suggesting that other regions are required for



complete rescue of high temperature growth. Indeed, *S. cerevisiae* *COX2* and *COX3* showed small but positive effects when the recombinants lacking them were compared to the wild-type *S. cerevisiae* mitotype (Fig. 3-2B). The differential heat sensitivity is unlikely to be caused by fitness defects since the recombinants grew normally at 22°C (Fig. S3-4A).

We also found that nearly all mtDNA recombinants did not exhibit 4°C respiratory growth; one strain (S87) derived from the *atp6Δ* cross (Fig. 3-2B) was an exception, but another strain with the same mitochondrial genotype did not grow. The 4°C recombinant phenotypes suggest that cold tolerance might require multiple *S. uvarum* alleles and potentially a different set of genes than those underlying heat tolerance.



**Figure 3-2. Mitochondria-encoded genes affect heat and cold tolerance.** (A) Hybrids with *S. cerevisiae* (sc) and *S. uvarum* (su) mtDNA differ in high- and low-temperature growth. Growth is after 5 days (22°C and 37°C) or 124 days (4°C). (B) Recombinant strains (rows) derived from mutant crosses (left) are clustered by genotype (middle). Wild-type *S. cerevisiae* (wt\_sc) and *S. uvarum* (wt\_su) mitotypes are at the bottom and the top, respectively. Allele identity is shown

for 12.6k orthologous single nucleotide markers (sc and sc-90, *S. cerevisiae*; su and su-90, *S. uvarum*; mixed, heterozygous or chimeric; white, no data) in the *S. cerevisiae* gene order (bottom). 37°C growth is the average size of non-petite colonies on glycerol plates (right). The presence of 4°C glycerol growth is indicated by solid squares (far right). (C) Effect size of *S. cerevisiae* alleles on 37°C growth on glycerol is shown, with error bars representing 95% confidence intervals. The y-axis is scaled to the phenotype of wild-type *S. uvarum* and *S. cerevisiae* mitotype (horizontal lines). tRNAs are labeled by their single letter amino acid code and a black bar. Blue dashed lines indicate genome positions of *S. uvarum* genes compared to *S. cerevisiae*.

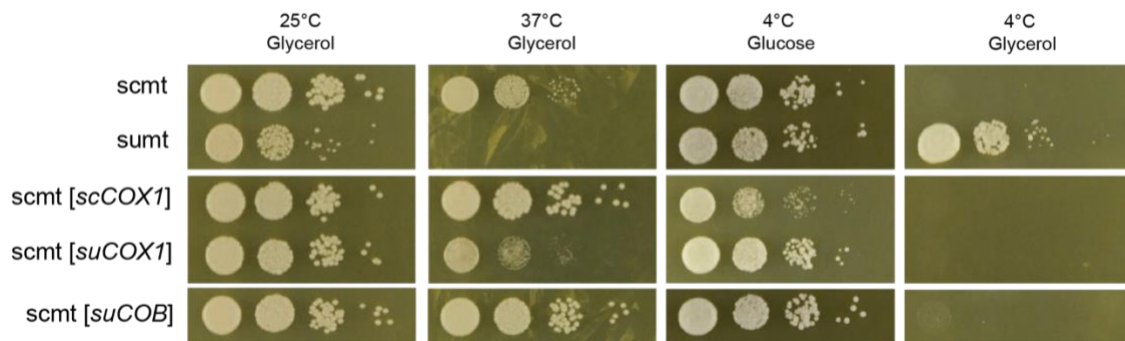
#### *COX1 protein divergence affects both thermotolerance and cryotolerance*

Because the recombinant strains did not resolve heat tolerance to a single gene, we tested individual genes by replacing *S. cerevisiae* with *S. uvarum* alleles via biolistic transformation (Bonney and Fox 2001) (Fig. S3-6). We obtained allele replacements for two of the four genes in the region conferring heat tolerance (Fig. 3-3). For both genes we used intronless alleles to eliminate incompatibilities in splicing (Chou et al. 2010).

We observed a significant difference between *S. cerevisiae* and *S. uvarum* *COX1* alleles for respiratory growth at 37°C in the hybrid background, with the *S. uvarum* allele being heat sensitive. The effect was not present at room temperature, and the *S. uvarum* allele conferred a growth advantage on glucose at 4°C. Thus, divergence in the *COX1* coding sequence (CDS) affects both heat and cold tolerance. However, *COX1* alleles do not explain the entire difference between the two species' mitotypes: the strain bearing *S. uvarum* *COX1* had an intermediate

level of heat tolerance and did not confer cold tolerance on glycerol, suggesting that other mitochondrial genes are involved. The moderate effect of the *COX1* alleles is also consistent with the small effect sizes shown by recombinant analysis (Fig. 3-2C). Surprisingly, the *COX1* allele difference is only seen in the hybrid and not in a diploid *S. cerevisiae* background (Fig. S3-7), suggesting that the allele difference in the hybrid depends on a dominant interaction with the *S. uvarum* nuclear genome.

The *S. uvarum* *COB* allele replacement rescued respiratory growth at high temperature, demonstrating that the *S. uvarum* *COB* protein is not heat sensitive. We were unable to generate the *S. cerevisiae* intronless *COB* allele replacement for comparison. Notably, both the intronless *S. cerevisiae* *COX1* and *S. uvarum* *COB* allele replacement strains exhibited better growth than wild-type *S. cerevisiae* mtDNA at 37°C (Fig. 3-3), implying a dominant-negative role of these introns in the hybrid at high temperature.



**Figure 3-3. *COX1* coding alleles affect growth at high and low temperature.** Hybrids carrying allele replacements and two wild-type controls were plated with 1:10 serial dilution and incubated at indicated temperatures. Growth is after 4 days for 25°C and 37°C, 25 days for 4°C on glucose, 53 days for 4°C on glycerol. sc, *S. cerevisiae*; su, *S. uvarum*; mt, mtDNA. Alleles in the brackets were integrated into their endogenous position in *S. cerevisiae* mtDNA.

## Discussion

In *Saccharomyces* species, the mitochondrial genome is not essential for viability, is large compared to insects and mammals (~86 kb), and is quite variable in intron content (Wolters et al. 2015; Wu et al. 2015). While the mitochondrial genome can recombine and introgress between species (Leducq et al. 2017; Peris, Arias, et al. 2017), it also contributes to reproductive isolation through incompatibilities with the nuclear genome (Lee et al. 2008; Chou et al. 2010; Spirek et al. 2015). Our results show that the mitochondrial genome also makes a significant contribution to one of the most distinct phenotypic differences among the *Saccharomyces* species: their thermal growth profile. Below, we discuss the implications of our results in relationship to the genetic architecture of species' phenotypic differences, the role of cyto-nuclear interactions in phenotypic evolution and reproductive isolation, and mitochondria as a hotspot in the evolution of *Saccharomyces* species.

### *Genetic architecture of interspecies differences in thermotolerance*

Crosses between closely related, inter-fertile species have shown that phenotypic divergence can be caused by a few loci of large effect, many loci of small effect or a mixture of the two (Orr 2001). In this study, we carried out a genome-wide non-complementation screen between two diverged yeast species. Out of 4,792 non-essential genes in our study, we found only one gene (*HFA1*) that showed a moderate effect on heat tolerance regardless of the mtDNA effect (Fig. 3-1C). Of relevance, 178 *S. cerevisiae* deletions are sensitive to 37°C (Auesukaree et al. 2009); a rate comparable to a subsample we examined in this study (78/2251). We can thus

conclude that the vast majority of the *S. uvarum* alleles tested exhibited no detectable loss of function at a temperature they do not experience in their native genome. However, our non-complementation screen had some limitations. We did not test essential genes and could not detect genes whose effects were masked by mtDNA inheritance or epistasis, which could occur due to the hybrid carrying an otherwise complete complement of both nuclear genomes.

We found allele differences in *HFAI* affect heat tolerance. *HFAI* encodes a mitochondrial acetyl-CoA carboxylase and participates in mitochondrial fatty acid synthesis, a process essential to cellular respiration and mitochondrial biogenesis (Kastaniotis et al. 2017). While disruption of *HFAI* in *S. cerevisiae* resulted in a low level of lipoic acid and consequently a temperature-dependent respiratory defect (Hoja et al. 2004; Suomi et al. 2014), the hemizygote with only the *S. uvarum* allele showed heat sensitive growth on glucose but not glycerol (Fig. S3-1C), suggesting that the divergence in heat tolerance of *HFAI* might not be directly linked to its role in respiration. Further investigation is needed to elucidate the molecular mechanism by which *HFAI* impacts thermal divergence.

Although our screen led us to discover a pronounced temperature dependent effect of mtDNA on respiratory growth and a more subtle effect on fermentative growth, the mtDNA effect explains only a small portion of the large difference in heat tolerance between the two species. The *S. cerevisiae* × *S. uvarum* hybrid without mtDNA grows at both 37°C and 4°C on glucose (Fig. 3-2B), indicating that the nuclear genomes carry dominant factors that remain to be identified.

Despite the small number of genes in the mitochondrial genome, our results show multiple genes within the mitochondrial genome influence heat tolerance. In addition to the large

effect of the *COX1-COB* region, recombinants that inherited *S. uvarum* *COX2* and/or *COX3* are considerably more heat sensitive than a hybrid with a complete *S. cerevisiae* mtDNA genome. Furthermore, while the *COX1*-linked region showed the largest effect, the *COX1* CDS does not explain the entire difference between two species' mitotypes. Although we ruled out protein-coding changes in *S. uvarum* *COB* to be heat sensitive, changes in the other protein-coding sequences and in gene expression remain to be tested.

The cause of mtDNA-mediated differences in cryotolerance is more opaque. At 4°C, only one recombinant with a significant fraction of *S. cerevisiae* mtDNA grew better than hybrids with an *S. cerevisiae* mitotype, suggesting that multiple *S. uvarum* alleles are required for cold tolerance. Although we showed that *S. uvarum* *COX1* increased cold tolerance on glucose, the effect is not seen on glycerol, suggesting its effect on respiration might depend on the presence of other *S. uvarum* mitochondrial alleles. However, because the recombinants were all isolated at 37°C, it is possible that they all share some other genetic element or change that facilitates heat tolerance but inhibits 4°C growth.

#### *Cyto-nuclear interactions in Saccharomyces evolution*

In addition to mitochondria-encoded genes, approximately 1,000 nuclear genes function in the mitochondria, many of which are involved in expression and regulation of mitochondrial genes and formation of the multi-subunit cytochrome b and c complexes (Vögtle et al. 2017). Among *Saccharomyces* species, multiple cyto-nuclear incompatibilities have been shown to contribute to reproductive isolation. *S. uvarum* *AEP2* cannot regulate the translation of *S. cerevisiae* *ATP9* mRNA (Lee et al. 2008), while *S. cerevisiae* *MRS1* cannot splice introns of *S. paradoxus* and *S. uvarum* *COX1* (Chou et al. 2010). Additionally, the *S. uvarum* RNA binding

protein *CCMI* has reduced affinity for the *S. cerevisiae* 15s rRNA (Jhuang et al. 2017). While these incompatibilities affect the construction of cybrids, where mtDNA from different species was introduced into *S. cerevisiae* (Spirek et al. 2015), the phenotypic consequences besides loss of respiration is not known.

Our results show that the mitochondrial genomes of *Saccharomyces* species influence both heat and cold tolerance and provide multiple lines of evidence for the role of cyto-nuclear interactions. First, the temperature effects of species' mitotypes interact with nuclear background (Fig. S3-1). While *S. cerevisiae* hybrids without mtDNA ( $\rho^0$ ) grow similarly on glucose medium, *S. cerevisiae* mtDNA confers different levels of heat tolerance in hybrids with *S. paradoxus*, *S. uvarum*, and *S. kudriavzevii*, the latter of which only grows slightly better than the  $\rho^0$  hybrid.

We also observed interactions between the *COXI* allele replacements and their nuclear background. *COXI* showed allele differences at high and low temperatures in the hybrid but not in *S. cerevisiae*. This difference can be explained by a species-specific dominant interaction, as might occur when there are hybrid protein complexes (Piatkowska et al. 2013). In this scenario, *S. uvarum* *COXI* can function with interacting *S. cerevisiae* proteins at high temperature but exhibits a loss of function when interacting with temperature sensitive *S. uvarum* nuclear factors that are dominant to their *S. cerevisiae* orthologs. The nuclear factor is unlikely to be the previously reported intron-splicing factor *MRS1* because our *COXI* alleles are intronless.

However, introns might affect temperature sensitivity. The intronless *S. cerevisiae* *COXI* and *S. uvarum* *COB* alleles showed better respiratory growth at 37°C than wild-type *S. cerevisiae* mtDNA, suggesting a dominant negative role of introns in the hybrid. In *Saccharomyces*, the



number and presence of mitochondrial introns is variable between species (Sulo et al. 2017). This contrasts with high conservation of mitochondrial protein coding sequences, which show over 90% sequence identity between *S. cerevisiae* and *S. uvarum*, much higher than the 80% average of nuclear-encoded genes (Kellis et al. 2003). The rapid evolution of introns might require co-evolution of splicing factors, such as *COX1* and *MRS1*. The wild-type hybrid with *S. cerevisiae* mtDNA might be under burden of intron splicing at high temperature caused by dominant negative *S. uvarum* splicing factors. Nevertheless, many introns self-splice and/or encode maturases or homing-endonucleases, which could be temperature sensitive in a nuclear-independent manner.

There is no clear indication that previously reported incompatibilities contribute to the mtDNA temperature phenotypes. The reported cyto-nuclear incompatibilities are recessive, and thus should not contribute to the hybrid phenotypes. For example, although the *S. cerevisiae* *MRS1* is incompatible with *S. uvarum* *COX1*, the latter can be correctly spliced by *S. uvarum* *MRS1* in the diploid hybrid, at least at permissive temperatures. One possibility is that *S. uvarum* *MRS1* is heat sensitive, which would explain the heat sensitivity of the *S. uvarum* mitotype because neither the *S. cerevisiae* nor *S. uvarum* *MRS1* would splice *S. uvarum* *COX1* at high temperature. Heat sensitivity of *S. uvarum* *MRS1* was tested in our non-complementation screen, but the result was inconclusive. The *S. cerevisiae* *MRS1* deletion was complemented by the *S. uvarum* allele in the *MATa* (BY4741) cross; but its effect was masked by mtDNA inheritance in the *MATα* (BY4742) cross. In this regard it is worth noting that *S. cerevisiae* chromosome 9, which carries *MRS1*, is duplicated in three of the recombinant strains; in two cases, these strains show increased 37°C growth compared to similar genotypes (Table S3-1).

## *Mitochondrial DNA and yeast evolution*

It has been proposed that mtDNA plays a disproportionate role in Dobzhansky-Muller incompatibilities. Although it is a small genome, it heavily interacts with nuclear genes and has a high nucleotide substitution rate, leading to co-evolution of the mitochondrial and nuclear genomes and multiple interspecific incompatibilities (Burton and Barreto 2012). Has adaptation played a role in driving these incompatibilities? Although no direct links are proven, evolution of the mitochondrial genome and mito-nuclear epistasis has been linked to multiple phenotypes (Solieri et al. 2008; Albertin et al. 2013; Picazo et al. 2014), including 37°C growth (Paliwal et al. 2014; Leducq et al. 2017; Wolters et al. 2018), and deficiencies in mitochondrial DNA cause heat sensitivity (Zubko and Zubko 2014). Here, we show that mtDNA is important for evolution of heat and cold tolerance in distantly related species, caused by the accumulation of multiple small-to-medium effect changes and potentially mito-nuclear epistasis. Taken together, the present and prior findings point to mtDNA as an evolutionary hotspot for yeast speciation and adaptation.

## **Materials and Methods**

### *Strains, growth conditions, and genetic manipulations*

Strains used in this study are listed in Table S3-1. *S. cerevisiae* was maintained on YPD (1% yeast extract, 2% peptone, 2% dextrose) at 30°C; *S. uvarum* and *S. cerevisiae* × *S. uvarum* hybrids were maintained on YPD at room temperature. Strains were also grown on complete medium (CM, 0.3% yeast nitrogen base with amino acids, 0.5% ammonium sulfate, 2% dextrose), or dropout medium (CM-xxx, 0.13% dropout powder, 0.17% yeast nitrogen base,

0.5% ammonium sulfate, 2% dextrose) where xxx represents the missing amino acids when appropriate. SDPSer medium (synthetic dextrose proline D-serine, 2% dextrose, 0.17% yeast nitrogen base without ammonium sulphate or amino acids, 5 mg/ml L-proline, 2 mg/ml D-serine) was used to select for *dsdAMX4* (Vorachek-Warren and McCusker 2004). Antibiotics were added to media when selecting for *KanMX*, *NatMX*, and *hphMX*. YPGly medium (1% yeast extract, 2% peptone, 3% glycerol) was used to examine respiratory growth.

*S. cerevisiae* and *S. uvarum* strains were mated by mixing strains with opposite mating types on YPD at room temperature overnight. Diploid hybrids were obtained by plating the mating mixture to double selection medium and confirmed by mating-type PCR.

Transformations in this study followed standard lithium acetate methods (Gietz et al. 1995). When transforming *S. uvarum* or *S. cerevisiae* × *S. uvarum* hybrid, we used 37°C for heat shock and room temperature for incubation.

Strains lacking mitochondrial DNA (*rho*<sup>0</sup>) were generated by overnight incubation with shaking in liquid minimal medium (MM, 0.17% yeast nitrogen base without amino acid and ammonium sulfate, 0.5% ammonium sulfate, 2% dextrose) containing 25 ug/ml ethidium bromide. Following incubation, the culture was plated to YPD and YPGly to identify non-respiring colonies.

#### *Interspecific hemizygote collections*

*trp1 S. uvarum* strains YJF2600 and YJF2601 were constructed by replacing *TRP1* with *hphMX4* in YJF1449 (*MATa*) and YJF1450 (*MATα*) in the CBS7001 background (Scannell et al. 2011), respectively. The haploid yeast deletion collections derived from BY4741 (*MATa his3ΔI*

*leu2Δ0 met15Δ0 ura3Δ0*) and BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) were arrayed in 384-well format using a Singer ROTOR (Singer Instruments, Watchet UK) and mated to *trp1 S. uvarum* strains. Diploids were selected on CM-trp-his-leu-lys-ura plates. The resulting two interspecific hybrid collections were hemizygous for 4,792 genes.

The hemizygote collections were screened for non-complementation using the following conditions: 1) YPD at room temperature, 30°C, 35°C and 37°C; 2) CM with 0.5 mM copper sulfate at room temperature; and 3) YPD with 10% ethanol at 30°C. Pictures of plates were taken on the second and fifth day of incubation using a Nikon D3100 camera. Colonies that were visually smaller than wild-type (represented by most of the hemizygotes on the same plate) on day 5 were scored as sensitive, ranging from no growth to slightly sensitive growth. For heat, copper, and ethanol stresses, respectively, we found 145, 137, and 26 non-complemented genes from the BY4741 (*MATa*) cross and 221, 134, and 19 from the BY4742 (*MATα*) cross, resulting in an intersection of 80, 13, and 2 genes.

Respiration-deficient strains (petites) were identified by plating the haploid deletion collection strains on YPGly at 30°C. To estimate the rate of temperature-sensitive deletions, we sampled six plates (~2.3k strains) from the haploid deletion collections and assayed their growth on YPD plates at room temperature and 37°C. The rate of heat-sensitive deletions in the subsample was 78/2251.

#### *Validation of non-complementing genes*

We first repeated the non-complementation test in another strain background. We made deletions of candidate genes (*HFA1* for heat; *TDA1*, *TDA9*, *GGC1*, *TDA4*, *RPL39*, *ADD66*, *YOL075C*, *CUP2*, and *CAJ1* for copper) by *KanMX* in an *S. cerevisiae* strain YJF173 in the same

way as the deletion collection, with the exception that the coding region of *HFA1* was defined according to (Suomi et al. 2014). The knockout strains were then crossed to a *S. uvarum rho*<sup>0</sup> strain (YJF2760). Phenotypes of the hemizygotes were assessed at the same conditions as in the screen, and only phenotypes of *HFA1* and *CUP2* were replicated.

Reciprocal hemizygotes were generated for *HFA1* and *CUP2*. Orthologs of *S. cerevisiae HFA1* and *CUP2* were knocked out in *S. uvarum* strain YJF1450 with *KanMX*. The orthologs were defined according to (Scannell et al. 2011); for *HFA1*, we included an extra 477 bp upstream of the ATG for the *S. uvarum* allele, based on translation from a non-AUG start codon at position -372 in *S. cerevisiae* (Suomi et al. 2014). The *S. uvarum* deletion strains were then crossed to *S. cerevisiae* (YJF173), and the resulting hemizygotes were genotyped by PCR and found to carry *S. cerevisiae* mtDNA. Phenotypes of the two reciprocal hemizygotes were assessed on the same plate, under the same conditions as in the screen.

#### *Interspecific hybrids with reciprocal mitotypes*

Interspecific hybrids with reciprocal mitotypes were generated by crossing a *rho*<sup>+</sup> strain from one species to a *rho*<sup>0</sup> strain from another species. Two *rho*<sup>0</sup> colonies from each strain were crossed to control for possible mutagenic effects of the ethidium bromide treatment. Mitotype was confirmed by PCR using primers targeting the tRNA clusters in mtDNA (forward 5'-CCATGTTCAAATCATGGAGAGA-3', reverse 5'-CGAACTCGCATTCAATGTTTGG-3'; 95°C 2min; 95°C 30s, 50°C 30s, 72°C 30s for 30 cycles; 72°C 5min). The expected product sizes are 167 bp for *S. cerevisiae*, 131 bp for *S. paradoxus*, 218 bp for *S. kudriavzevii*, and 100 bp for *S. uvarum*.

### *Crosses with mitochondrial knockouts*

*S. uvarum* strain YJF2600 (*MATa hoΔ::NatMX trp1Δ::hphMX4*) and YJF2601 (*MATα hoΔ::NatMX trp1Δ::hphMX4*) were crossed to previously constructed *S. cerevisiae* mitochondrial knockout strains (Steele et al. 1996; Bonnefoy and Fox 2000; Perez-Martinez et al. 2003; Rak et al. 2007; Ding et al. 2009; Rak and Tzagoloff 2009). *S. cerevisiae* strains with wild-type mtDNA were crossed in parallel as control. *MATa* and *MATα* strains were mixed on YPD and incubated at room temperature overnight. The mating mixtures were either replica-plated (initial trial) or resuspended in sterile water and plated (second trial) onto YPGly. The YPGly plates were incubated at 37°C for 7-10 days to select for 37°C-respiring recombinants. The mating mixtures of *cox2Δ* and *cox3Δ* crosses were also plated to CM-trp-his-leu-lys-ura at room temperature to select for diploid hybrids, which allowed us to estimate the recombination rate to be around 0.05-0.1%. 37°C-respiring colonies were picked and streaked on YPD at room temperature for single colonies. For the initial trial, the 37°C-respiring cells were streaked on YPD twice. For the *cox1Δ* and *atp6Δ* crosses, the plates were left at room temperature for 3 days after 7 days at 37°C incubation and colonies growing from the recovery period were also picked and streaked. We also tried selecting for recombinants at 33°C and 35°C for the crosses with *cobΔ*, *atp6Δ* and *atp8Δ* strains, from which we isolated few recombinants at 37°C. However, selection at 35°C did not significantly increase either the number or the size of the recombinant colonies compared to 37°C, and 33°C is too low a temperature to distinguish any heat-tolerant recombinants from non-recombinant *S. uvarum* mtDNA; we thus did not sequence colonies from these selections. As a result, 3+12, 4+48, 3+25, 2+3, 0+7, and 0+1 strains (initial trial + second trial) from the *cox2Δ*, *cox3Δ*, *cox1Δ*, *cobΔ*, *atp6Δ* and wild-type D273-10B control crosses,

respectively, were generated. The total of 102 strains were subjected to whole genome sequencing and phenotyping.

#### *Spontaneous mitochondrial recombinants*

*S. cerevisiae* (YJF153, *MAT $\alpha$  ho $\Delta$ ::dsdAMX4*, YPS163 derivative) and *S. uvarum* (YJF1450, *MAT $\alpha$  ho $\Delta$ ::NatMX*, CBS7001 derivative) were mated and streaked onto SDPSer + clonNAT medium to select for diploid hybrids. 384 colonies on the double selection plates were picked and arrayed onto one YPD agar plate and subsequently pinned to YPD and YPGly and incubated at room temperature, 37°C and 4°C. Colony sizes on each plate were scored both manually and quantitatively using ImageJ (Rasband). Strains with recombinant-like temperature phenotypes (r114, r194, r262, r334, r347 and b2), along with two control strains (r21, r23) with typical phenotypes for *S. cerevisiae* and *S. uvarum* mitotypes, respectively, were subjected to whole genome sequencing and phenotyping.

#### *DNA extraction, library preparation, and sequencing*

For the unselected putative recombinants and their controls (r21, r23, b2, r334, r114, r194, r262, and r347), DNA was extracted using an mtDNA-enriching protocol (see below). For other strains sequenced in this study, genomic DNA was extracted from 22°C YPD overnight cultures inoculated with cells pre-grown on YPGly plates (ZR Fungal/Bacterial DNA MicroPrep kit, Zymo Research).

mtDNA was enriched following a protocol adapted from (Fritsch et al. 2014) and (Wolters et al. 2015). 50ml YPEG (1% yeast extract, 2% peptone, 2% ethanol, 2% glycerol) medium was inoculated with overnight YPD starter cultures, shaken at 300 rpm at 22°C. The

culture was collected at late-log phase (3,000g for 1 min) and the cell pellet was washed twice in 1ml sterile distilled water. The cells were then washed in buffer (1.2M Sorbitol, 50mM Tris pH 7.4, 50mM EDTA, 2% beta-mercaptoethanol) and centrifuged at 14,000 rpm for 3 minutes. The cell pellet was weighed, resuspended in Solution A (0.5M Sorbitol, 50mM Tris pH 7.4, 10mM EDTA, 2% beta-mercaptoethanol, 7ml/g wet weight cells) containing 0.2mg/ml Zymolyase (Zymo Research), and incubated at 37°C, 100 rpm for 45min for osmotic lysis. The suspension was then centrifuged at 4,000rpm for 10min. The supernatant was decanted to a new tube and centrifuged at 14,000 rpm for 15min, to get the crude mitochondrial pellet. The pellet was then incubated in DNase treatment solution [0.3M Sucrose, 5mM MgCl<sub>2</sub>, 50mM Tris-HCl pH 8.0, 10mM CaCl<sub>2</sub>, 100U/ml RQ1 DNase (Promega), 500ul/g initial wet weight] at 37°C, 100 rpm for 30min to remove nuclear DNA. 0.5M EDTA (pH 8.0) was added to a final concentration of 0.2M to stop the reaction. The mitochondrial pellet was then washed three times by repeated cycles of centrifugation at 15,000 rpm for 10min and resuspension in 1ml solution A to remove DNase, and then resuspended in 400ul Solution B (100mM NaCl, 10mM EDTA, 50mM Tris pH 8) and incubated at room temperature for 30min for lysis. mtDNA was isolated from the solution by phenol-chloroform extraction and ethanol precipitation, followed by a clean-up with DNA clean and concentrator -5 kit (Zymo Research). Alternatively, two samples (r21 and r262) were extracted with ZR Fungal/Bacterial DNA MicroPrep Kit (Zymo Research) by adding the Fungal/Bacterial DNA binding buffer to the lysed mitochondrial fraction and following the rest of the manufacturer protocol. The yield was typically 10-20ng/g wet weight cells and provided 10- to 100-fold enrichment of mitochondrial reads.

Paired-end libraries were prepared with Nextera DNA Library Preparation Kit (Illumina) with a modified protocol. Briefly, 3-5 ng DNA was used for each sample and the tagmentation



reaction was performed at a ratio of 0.25ul tagmentation enzyme/ng DNA. The tagmented DNA was amplified by KAPA HiFi DNA polymerase for 13 cycles (72°C 3min; 98°C 5min; 98°C 10s, 63°C 30s, 72°C 30s for 13 cycles; 72°C 5min). The PCR reaction was then purified with AMPure Beads. Paired-end 2x150 Illumina sequencing was performed on a MiniSeq by the DNA Sequencing Innovation Lab in the Center for Genome Sciences and System Biology at Washington University. 96 recombinants generated in the second trial of the mitochondrial mutant crosses were subsequently re-sequenced on a NextSeq 500 at Duke Center for Genomic and Computational Biology for deeper coverage. The NextSeq reads and MiniSeq reads were combined in the analysis. The reads were deposited at the Sequence Read Archive under accession no. SRP155764.

#### *Mitochondrial genome assembly*

The *S. uvarum* mitochondrial genome was assembled from high-coverage sequencing of r23. Before assembly, we confirmed that it carried a non-recombinant *S. uvarum* mitochondrial genome by mapping the reads to CBS380 (Okuno et al. 2016), an *S. eubayanus* × *S. uvarum* × *S. cerevisiae* hybrid that inherited the mitochondria from *S. uvarum*. To assemble the mitochondrial genome, reads were first cleaned with trimmomatic (Bolger et al. 2014) to remove adapters. They were then assembled using SPAdes assembler (Bankevich et al. 2012), included in the wrapper iWGS (Zhou et al. 2016), to produce contigs. Contigs were scaffolded to produce the final assembly through comparison with the output assembly of MITObim (Hahn et al. 2013). The assembly was annotated with MFannot Tool (<http://megasun.bch.umontreal.ca/RNAweasel>); *ORF1* (F-*SceIII*) annotation was added manually using Geneious R6 (Kearse et al. 2012). The assembled r23 mitochondrial genome is 64,682 bp and has a total of 5,874 gapped bases

(GenBank accession no. MH718505). Most gaps are in the intergenic regions, one gap is in *VARI*, and 3 small gaps are in the introns of *COB*. The r23 mitochondrial genome is 99% identical to CBS380, based on BLAST results.

### *Read mapping and allele assignment of recombinants*

Illumina reads were mapped to a reference that combined the mitochondrial genomes of *S. cerevisiae* (S288C-R64-2-1) and *S. uvarum* (r23 mitochondria assembled in this study), using end-to-end alignment in Bowtie 2 (Langmead and Salzberg 2012). Duplicated reads and reads with high secondary alignment scores ( $XS \geq AS$ ) or low mapping quality ( $MQ < 10$ ) were filtered out. Using this method, reads from hybrids with non-recombinant *S. cerevisiae* or *S. uvarum* mtDNA were >99.9% correctly mapped to their reference genomes (49,496/49,504 for *S. cerevisiae*, 161,712/161,714 for *S. uvarum*). To characterize aneuploidy and the ratio of mitochondrial to nuclear reads, the reads were re-mapped to a reference file combining *S. cerevisiae* (S288C-R64-2-1) and *S. uvarum* (Scannell et al. 2011) reference genomes, using the same method. Coverage of nucleotide positions and of chromosomes was generated by *samtools depth* and *samtools idxstats*, respectively.

For data visualization and identification of recombination breakpoints, we assigned allele identity for each nucleotide in orthologous regions in the two reference mitochondrial genomes. The total length of orthologous sequences is 16.5kb (nucmer alignment) and contains mostly coding and tRNA sequences. After removing sites with no coverage in control strains, 12.6k nucleotide positions were subjected to data visualization and allele calling. We called the allele identity of a given nucleotide position based on the ratio of reads that mapped to the *S. cerevisiae* reference allele to the total number of reads that mapped to the two orthologous alleles

( $rsc = sc / (sc + su)$ ): rsc of 1 (or no lower than the non-recombinant *S. cerevisiae* mtDNA control) was called *S. cerevisiae*, rsc of 0 (or no higher than the non-recombinant *S. uvarum* mtDNA control) was called *S. uvarum*,  $rsc > 0$  and  $< 1$  were called mixed. Sites without coverage of either allele were treated as missing data. A relaxed threshold was used in data visualization to account for noise in read mapping ( $rsc > 0.9$  was called *S. cerevisiae*, labeled as “sc-90”;  $rsc < 0.1$  was called *S. uvarum*, labeled as “su-90”). Using this method, a total of 90 sequenced strains were confirmed to be recombinants.

For quantifying the effect size of *S. cerevisiae* alleles, we counted the number of reads mapped to each protein-coding gene, tRNA and rRNA by htseq-count. For each gene, we tested the allele effect across 90 recombinants using a linear model:  $phenotype \sim allele + petite$ , where *allele* is the ratio of *S. cerevisiae* reads for a given gene and *petite* is the empirically determined petite rates (see below). Because we used the ratio of *S. cerevisiae* reads to represent allele identity, the model does not assume dominance; a heterozygous individual (i.e. read ratio = 0.5) should have an intermediate phenotype. P-values were extracted from the models and adjusted by the false discovery rate (Benjamini & Hochberg method) to correct for multiple comparisons. While the p-value for the *petite* term is significant in some models, its effect was always estimated to be positive. Because high petite rates should lead to small colonies, we do not consider petite rate to significantly contribute to the phenotype. Additionally, aneuploidy and mtDNA copy number variation were present in several recombinants, but the addition of the two variables to the model did not change the effect size and significance of the *allele* term ( $phenotype \sim allele + petite + aneuploidy + copy$ , where *aneuploidy* is a binary variable indicating presence/absence of chromosomal duplication and *copy* is the ratio of mitochondrial to nuclear reads).

The unselected putative recombinants were sequenced to high coverage, so we generated contigs and assemblies as in “Mitochondrial genome assembly”. The contigs were mapped to *S. cerevisiae* (r21) and *S. uvarum* (r23/CBS380/CBS7001) assemblies in Geneious R6 to identify the breakpoints. For the recombinants of lower quality assemblies (r194, r347, and b2), the contigs were mapped to the best recombinant assembly r114 to improve recombinant construction. Results were confirmed by retaining the Illumina reads from the mitochondrial genome, using both reference mitochondrial genomes as baits in HybPiper (Johnson et al. 2016) and mapping them to the reference mitochondrial genomes using Geneious R6.

### *Recombinant phenotypes*

Recombinant strains were first grown on YPGly plates to enrich for respiring cells, then in liquid YPD shaken at room temperature overnight. The overnight culture was diluted 1:10<sup>5</sup>, spread on YPD and YPGly plates and incubated at 22°C, 37°C, or 4°C. Pictures of plates were taken on the 5<sup>th</sup> day for 22°C and 37°C YPD plates, on the 6<sup>th</sup> day for 22°C and 37°C YPGly plates and on the 68<sup>th</sup> day for 4°C YPD and YPGly plates. Colony sizes on YPGly plates were acquired by the *Analyze Particles* function in ImageJ (Rasband). Non-single colonies were filtered out both by manually marking problematic colonies during analysis and by roundness threshold (roundness > 0.8 for non-petite colonies). For each strain, sizes of all the non-petite colonies (colony size > 200 units) were averaged; if no cells were respiring at a given condition, the average of all the (micro)colonies was used instead. Petite rates of the overnight cultures were recorded by counting big/small colonies on 22°C YPD and normal/micro colonies on 22°C YPGly plates, and the two values were averaged. Control strains carrying wild-type *S. cerevisiae* or *S. uvarum* mtDNA in the background of D273-10B × CBS7001 were phenotyped in parallel.

Initially the ~90 strains were phenotyped in three batches. We accounted for the batch effect for the 37°C data by picking 3-4 strains from each batch and repeating the phenotyping process on the same day at 37°C. Linear models between old data and new data were generated for each batch separately and were used to adjust for an overall batch effect. The 22°C colony sizes were not adjusted.

### *Mitochondrial allele replacement*

Mitochondrial transformation was performed as previously described (Bonney and Fox 2001) (Fig. S3-6). Intronless mitochondrial alleles were synthesized by Biomatik. The alleles were Gibson-assembled into an *ARG8m*-bearing pBluescript plasmid, such that the mitochondrial allele is flanked by 69 bp and 1113 bp *ARG8m* sequences at its 5' and 3' end, respectively (Fig. S3-6C). Sequences of the assembled plasmid were confirmed by Sanger sequencing.

Mitochondrial knockout strains were first transformed with *P<sub>GAL</sub>-HO* to switch mating types and validated by mating type PCR. In these strains, the target gene was replaced with *ARG8m*, so our constructs carrying the allele of interest can integrate into their endogenous loci by homologous recombination with *ARG8m* (Fig. S3-6C).

We bombarded the mitochondrial plasmid and pRS315 (CEN plasmid carrying *LEU2*) into *S. cerevisiae* strain DFS160 (*MAT $\alpha$  ade2-101 leu2 $\Delta$  ura3-52 arg8 $\Delta$ ::URA3 kar1-1, rho<sup>0</sup>*) (Steele et al. 1996) using a biolistic PDS-1000/He particle delivery system (Bio-Rad) and selected for Leu<sup>+</sup> colonies on MM plates. The colonies were replica-mated to the mitochondrial knockout strains at 30°C for 2 days. The mating mixtures were replica-plated to YPGly plates and incubated at 30°C. YPGly<sup>+</sup> colonies were streaked on YPD and mating types were determined by PCR. We also isolated the DFS160-derived parent strains that gave rise to the

YPGly<sup>+</sup> colonies from the master plates. For *S. cerevisiae* *COX1* and *COB* alleles, the parent strains were re-mated to the knockout strains for confirmation.

The YPGly<sup>+</sup> colonies carry a mitochondrial genome with the allele of interest integrated at their endogenous loci. Because of the *kar1-1* mutation in DFS160, we were able to isolate YPGly<sup>+</sup> colonies that are diploid, *MATa* haploid, or *MATα* haploid. We crossed the *MATa* transformant (D273-10B background) to an *S. uvarum* rho<sup>0</sup> strain (YJF2760). The hybrid strain and the diploid *S. cerevisiae* strains directly obtained from the mitochondrial transformation were phenotyped at room temperature, 37°C, and 4°C on YPD and YPGly by spot dilution assays. The allele identity of all the phenotyped strains was confirmed by PCR and restriction digest.

**Chapter 4: Multiple changes underlie allelic  
divergence of *CUP2* between *Saccharomyces* species.**

## Abstract

Under the model of micromutationism, phenotypic divergence between species is caused by accumulation of many small-effect changes. While mapping the causal changes to single nucleotide resolution could be difficult for diverged species, genetic dissection via chimeric constructs allows us to evaluate whether a large-effect gene is composed of many small-effect nucleotide changes. In the non-complementation screen described in Chapter 3, I found allele difference of *CUP2*, a copper-binding transcription factor, underlie divergence in copper resistance between *S. cerevisiae* and *S. uvarum*. Here, I tested whether the allele effect of *CUP2* was caused by multiple nucleotide changes. By analyzing chimeric constructs containing four separate regions in the *CUP2* gene, including its distal promoter, proximal promoter, DNA binding domain and transcriptional activation domain, I found that all four regions of the *S. cerevisiae* allele conferred copper resistance, with the proximal promoter showing the largest effect, and that both additive and epistatic effects are likely involved. These findings support the model of micromutationism and suggest an important role of both protein coding and *cis*-regulatory changes in evolution.



## Introduction

Our view of the genetic architecture of evolution could be biased by the resolution of genetic analysis. While many mapping studies identified large-effect genes underlying evolution, they are not necessarily caused by large-effect mutations (Rockman 2011): over a long evolutionary timescale, multiple small-effect nucleotide changes could accumulate in one gene, as suggested by the model of micromutationism. This view is supported by recent dissection of large-effect loci (McGregor et al. 2007; Frankel et al. 2011; Engle and Fay 2012) (reviewed in (Martin and Orgogozo 2013)). However, there are also examples of a few nucleotide changes causing large phenotypic effects, e.g. (Prud'homme et al. 2006; Nagy et al. 2018). A complete understanding of the genetic architecture of species divergence should include more intragenic fine-mapping studies, which will also provide insight on the relative contribution of *cis*-regulatory and coding changes and the prevalence of epistasis in evolution.

Using a genome-wide non-complementation screen, we previously found that divergence of *CUP2* contributed to evolution of copper resistance in *Saccharomyces* species. *S. cerevisiae* can tolerate high concentration of copper sulfate, a stress associated with vineyard environment. Although the level of copper resistance is variable among *S. cerevisiae* strains (Fay et al. 2004; Kvitek et al. 2008), its relatives, *S. paradoxus* and *S. uvarum*, are usually copper sensitive (Kvitek et al. 2008; Dashko et al. 2016). Through non-complementation and reciprocal hemizyosity test, we found that the *S. cerevisiae* *CUP2* allele conferred higher copper resistance compared to the *S. uvarum* allele. *CUP2* encodes a copper-binding transcription factor and regulates Cup1p, a major copper-activated metallothioneine in yeast (Buchman et al. 1989). Previous studies showed that *CUP2* is essential for *S. cerevisiae*'s copper resistance (Thiele

1988; Welch et al. 1989; Jin et al. 2008) and was involved in intraspecific variation (Kim et al. 2012). The sequences of *S. cerevisiae* and *S. uvarum* *CUP2* are substantially diverged (71.1% identical), making it difficult to map individual nucleotide substitutions. Thus, we dissected the effect of *CUP2* allele using chimeric constructs. We found that copper-resistant nucleotide changes are distributed throughout the gene, with *cis*-regulatory changes having a larger effect than the coding changes.

## Materials and Methods

*S. cerevisiae* strains in the S288C background and *S. uvarum* strains in the CBS7001 background (Scannell et al. 2011) were used in this study. The *S. uvarum* genome sequence and annotations were from Scannell et al. (2011). *CUP2* was knocked out with *KanMX4* in *S. cerevisiae* (YJF173, *MAT $\alpha$  ho-ura3-52*) and *S. uvarum* (YJF1450, *MAT $\alpha$  ho $\Delta$ ::NatMX*), respectively. Transformations in this study followed a standard lithium acetate procedure (Gietz et al. 1995), with the modification that room temperature and 37°C was used for incubation and heat shock of *S. uvarum*, respectively. Unless otherwise noted, *S. cerevisiae* was maintained at 30°C on YPD (1% yeast extract, 2% peptone and 2% dextrose) while *S. uvarum* and *S. cerevisiae*  $\times$  *S. uvarum* hybrids were maintained at room temperature.

Chimeric constructs were generated by Gibson assembly (Gibson et al. 2009). Promoters were defined from the end of the 5' ORF (*PMRI*) to the start codon of *CUP2*. CDS was defined from the start codon of *CUP2* to the stop codon, and our constructs also included the 3' non-coding region. To further dissect the effects of the promoter and CDS, the promoter was split at nucleotide position -291 for *S. cerevisiae* and its homologous position at -283 for *S. uvarum*. The CDS was split at position +367 for both alleles, based on the previously defined DNA binding

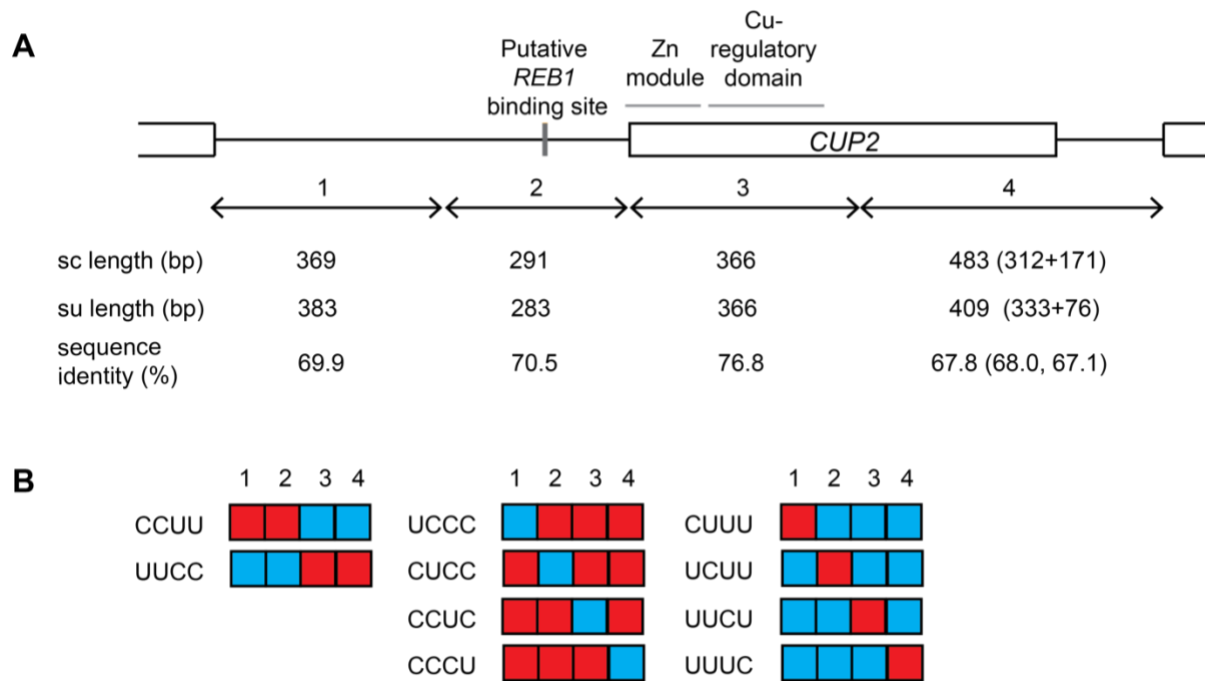
domain and transactivation domain (Buchman et al. 1989) (Fig. 4-1A). All positions are relative to the start codon of *CUP2*.

Segments of *CUP2* were PCR-amplified from *S. cerevisiae* or *S. uvarum* genomic DNA with Q5 polymerase (New England Biolabs). Promoter and CDS segments from different species were Gibson-assembled into pRS306 to generate promoter-swaps. Full-length *S. cerevisiae* and *S. uvarum* *CUP2* alleles were assembled in parallel for controls. An *S. cerevisiae* oak allele was included for comparison, which was amplified from genomic DNA of YJF153 (*MAT $\alpha$  ho $\Delta$ ::dsdAMX*), a YPS163 derivative. To split the promoter or CDS, the segments of interest were assembled into pRS306-derived plasmids pXL07 or pXL05, which respectively carry the full-length *S. cerevisiae* or *S. uvarum* allele, to replace the segment in the original allele. All constructs were Sanger-sequenced; one of the chimeras (CCUC) carried a deletion of A in a stretch of 14As in the *S. cerevisiae* promoter, but it did not seem to cause deleterious effects in the phenotypic assays.

The plasmids were linearized with BstBI (*CUP2* constructs) or StuI (vector control) and integrated into the *ura3* locus of *S. cerevisiae* *CUP2* knockout strain YJF2872 (*MAT $\alpha$  ho- ura3-52 cup2 $\Delta$ ::KanMX4*). The integrated strains were backcrossed to an *S. cerevisiae* strain YJF175 (*MAT $\alpha$  ho- ura3-52*) and sporulated to remove any second-site mutations. The resulting haploid *S. cerevisiae* strains carrying the *CUP2* deletion and chimeric constructs were then crossed to an *S. uvarum* *CUP2* knockout YJF2917 (*MAT $\alpha$  ho $\Delta$ ::NatMX cup2 $\Delta$ ::KanMX4*). The final interspecific hybrid was null for both *S. cerevisiae* and *S. uvarum* alleles at their endogenous loci and carried chimeric or full-length constructs at the *ura3* locus. The hybrids were genotyped by PCR and found to carry *S. cerevisiae* mitochondrial DNA.

Growth curves in copper-supplemented media were recorded by a BioTek microplate reader. Three biological replicates were used for each strain. Overnight cultures were 1:100 diluted into 200 ul complete media (CM, 0.3% yeast nitrogen base with amino acids, 0.5% ammonium sulfate, 2% dextrose) supplemented with 0, 0.2 or 0.5mM copper sulfate in a 96-well plate. The plate was incubated at room temperature (25-26°C), with the optical density (OD) at 600nm taken every 10min for 40h. The plate was shaken for 20s before each OD reading. To quantify growth differences, area under curve (AUC) was measured as the integral of the spline fit of growth curves using the grofit package (Kahm et al. 2010) in R. Copper resistance was represented by normalized AUC (nAUC), the AUC of copper treatments divided by the mean of AUC of the same strain under CM treatments.

Linear models were used to analyze the effects of each region. Data of the oak allele and the vector control were excluded in the models. The sum of nAUC across the two concentrations (snAUC) was used to represent copper resistance of each strain. The data were fit to two models: 1)  $snAUC \sim R1 + R2 + R3 + R4$ , to analyze the additive effects of region 1 to 4 ( $R1$  to  $R4$ ); 2)  $snAUC \sim (R1 + R2 + R3 + R4)^2$ , to analyze both additive and epistatic effects.  $R1$  to  $R4$  were categorical variables (C or U). P-values were extracted from the models and were adjusted by false discovery rate (Benjamini and Hochberg method) to correct for multiple comparisons.



**Figure 4-1. Design of *CUP2* chimeras.** **A.** Diagram of *CUP2* gene, with black lines representing non-coding regions and boxes representing coding regions. The alleles were split into 4 regions (1-4). Region 2 contains a putative *REB1* binding site and region 3 contains the DNA binding domain (DBD) (Buchman et al. 1989), including a 40-residue zinc module (Turner et al. 1998) and a ~60 residue copper regulatory domain (Graden et al. 1996). The diagram is drawn to scale of the *S. cerevisiae* allele, with the length of *S. cerevisiae* (sc) and *S. uvarum* (su) regions indicated below. Sequence identity is based on MUSCLE alignments, without counting gaps. Region 4 includes the 3' half of the coding sequence and the 3' intergenic sequence, of which the sequence length and identity was separately indicated in the parentheses. **B.** *S. cerevisiae* (C, red) and *S. uvarum* (U, blue) segments were assembled into 10 chimeric constructs, including promoter-swaps (left), different *S. uvarum* regions inserted into the *S. cerevisiae* allele (middle), and different *S. cerevisiae* regions inserted into the *S. uvarum* allele (right).

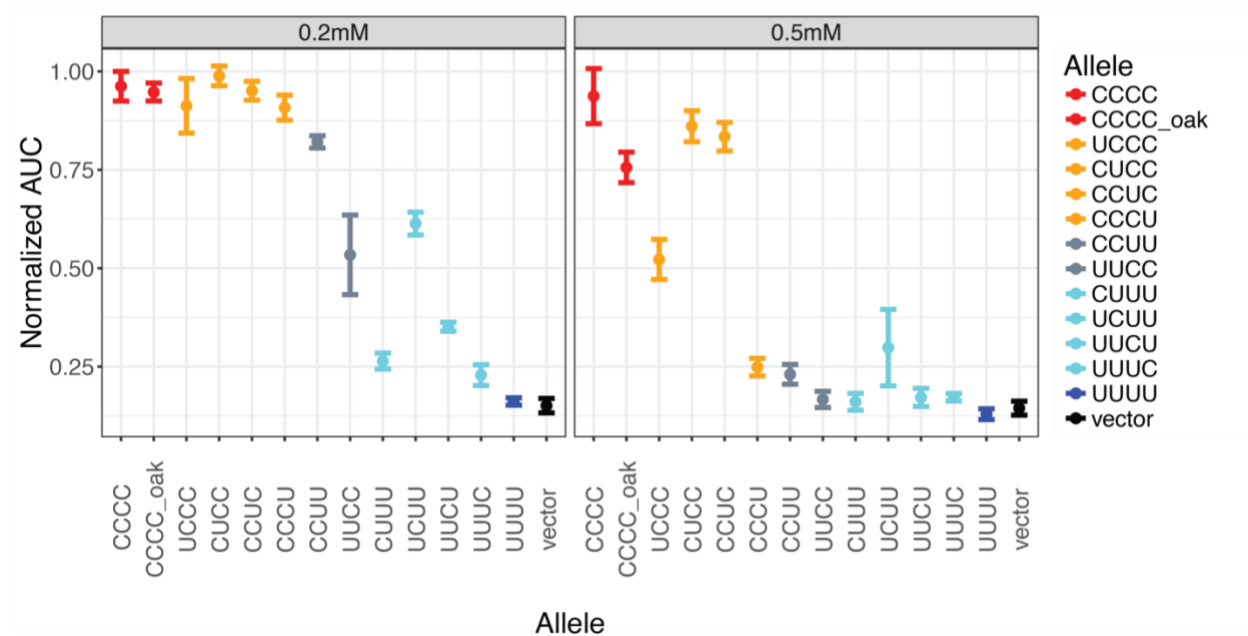
## Results

*S. cerevisiae* *CUP2* was shown to confer higher copper resistance than the *S. uvarum* allele by a reciprocal hemizyosity test. The two alleles share 71.1% sequence identity, with hundreds of nucleotide substitutions across the coding and non-coding regions. To understand whether the allele differences result from multiple nucleotide changes and the relative contribution of coding and *cis*-regulatory changes, we generated chimeric constructs between *S. cerevisiae* and *S. uvarum* *CUP2* alleles (Fig. 4-1) and integrated them into the *ura3* locus in *S. cerevisiae*. Copper resistance of the constructs was measured in a hybrid of *S. cerevisiae* and *S. uvarum*, in which the endogenous *CUP2* alleles were knocked out. The hybrid background was used in accordance with the reciprocal hemizyosity test, and the effects of chimeras were the same in *S. cerevisiae* (Fig. S4-1). The resistance of chimeras generally correlated with the number of *S. cerevisiae* segments in the constructs (Fig. 4-2). The *S. cerevisiae* promoter conferred higher resistance than the CDS (grey). The chimeras that split within the promoter or CDS regions further mapped the largest effect to the 3' half of the *S. cerevisiae* promoter (the UCUU construct), while the other three *S. cerevisiae* regions tested also conferred low-to-moderate levels of resistance when inserted into the *S. uvarum* allele (light blue, left panel), suggesting that multiple nucleotide changes underlie the allele effect of *CUP2*. While the combination of any three *S. cerevisiae* segments was sufficient to confer resistance to the 0.2mM copper treatment (orange), these chimeras showed various levels of sensitivity to 0.5mM, also consistent with a model of multiple changes.

The effect sizes of individual regions depended on the context, suggesting epistasis. For example, the *S. cerevisiae* distal promoter (region 1) and the C-terminal coding sequence (region

4) showed small effects when inserted into the *S. uvarum* allele (CUUU and UUUC, light blue), but replacing the *S. cerevisiae* region 1 and 4 with *S. uvarum* greatly reduced copper resistance (UCCC and CCCU, orange).

We also included a full-length *CUP2* allele from a copper-sensitive *S. cerevisiae* oak isolate for comparison. The oak allele has 12 nucleotide differences from the S288C allele used in the chimeras. While the oak allele showed a modest sensitivity to 0.5mM copper treatment, it was much more resistant than the *S. uvarum* allele (Fig. 4-2, right panel), suggesting that *S. cerevisiae* *CUP2* might have acquired many copper-resistant changes since the divergence with *S. uvarum*.



**Figure 4-2. Copper resistance of chimeric constructs.** *S. cerevisiae* × *S. uvarum* hybrids carrying the chimeric constructs were grown in labeled copper concentrations and their resistance was measured by area under curve (AUC) of OD<sub>600</sub> growth curves, normalized to their

growth in complete media. Points represent the mean of three biological replicates and error bars represent 95% confidence interval. The colors are based on the number of *S. cerevisiae* segments in the chimeras (red = 4, orange = 3, grey = 2, light blue = 1, blue or black = 0).

We used linear models to quantify the effect sizes and test for possible epistatic interactions (Table 4-1). We found that the model accounting for epistatic effects explained the data better than the model with only additive effects (0.974 vs. 0.839 for adjusted R-squared,  $p=1.94E-10$  in ANOVA). In the epistatic model, all four *S. cerevisiae* regions had significant effects on copper resistance, with region 2 showing the largest effect. Positive epistasis was detected between region 1 and 4, which corresponded to the high resistance of constructs CUCC and CCUC (Fig. 4-2, right panel). Region 1-2 and 2-3 showed modest negative interactions. These findings suggest that both changes with additive and epistatic effects contributed to the divergence of *CUP2* alleles.

Table 4-1. Additive and epistatic effects of *S. cerevisiae* *CUP2* regions on copper resistance.

Region*	Additive model		Epistatic model	
	Effect size	P-value†	Effect size	P-value†
(Intercept)	0.138	0.0841	0.197	0.000445
1	0.479	3.11E-06	0.314	9.37E-05
2	0.515	1.33E-06	0.801	1.29E-11
3	0.274	0.00267	0.333	5.59E-05
4	0.527	1.33E-06	0.211	0.00369
1*2			-0.339	0.000292
1*3			0.0754	0.370
1*4			0.594	1.47E-07



2*3			-0.232	0.00679
2*4			NA	NA
3*4			0.0381	0.622

\* Regions were defined as in Fig. 4-1A. The asterisks indicate interactions.

† P-values were adjusted by the false discovery rate (Benjamini and Hochberg method).

## Discussion

Evolution can occur through accumulation of many small-effect changes, but mapping small-effect changes can be technically challenging (Orr 2001). In the present study, we tested whether a relatively large effect on copper resistance caused by *CUP2* allele divergence is a consequence of multiple nucleotide changes. By splitting the *CUP2* gene into four regions and measuring their effects via chimeric constructs, we found that the *CUP2* allele difference was caused by accumulation of multiple small-to-medium effect changes, with the proximal promoter region showing the largest effect.

### *Multiple changes with small effects*

Our findings support the micromutationism view that evolution involves many small-effect changes. All four regions tested conferred copper resistance with various effect sizes, suggesting that the copper-resistant nucleotide substitutions are distributed throughout the *CUP2* gene. The largest effect was mapped to the proximal promoter. The promoter effect was unlikely to be caused by changes in transcription factor binding sites: there was only one putative *REB1* binding site in the *CUP2* promoter (YetFasCo database, (De Boer and Hughes 2012), Fig. 4-1A), and it is conserved across *Saccharomyces* species. The large effect of *CUP2* promoter supports

the previously suggested prominent role of *cis*-regulatory changes in long-term evolution (Stern and Orgogozo 2008). While *cis*-regulatory changes were often found to underlie morphological evolution, the example of *CUP2* along with several prior studies demonstrated that they are also important to physiological traits in yeast (Gerke et al. 2009; Engle and Fay 2012; Roop et al. 2016).

Cup2p consists of an N-terminal DNA binding domain (region 3) and a C-terminal transcriptional activation domain (region 4) (Buchman et al. 1989), with the former being more conserved (Fig. 4-1A). We found that the DNA binding domain of *S. cerevisiae* conferred moderate copper resistance when inserted into the *S. uvarum* allele. The gain of copper resistance could be due to changes in binding affinity to the *CUP1* promoter. The N-terminal of Cup2p is suggested to bind DNA via a zinc module and a copper-regulatory domain (Graden et al. 1996) (Fig. 4-1A), both of which contain amino acid differences between the two species. Further dissection of this region would help understand the molecular mechanism of *CUP2*-mediated copper resistance. However, these dissections are expected to become increasingly difficult under the micromutational model.

While all four regions showed different levels of additive effects, the context-dependent effect sizes of individual regions suggest epistasis. The *S. cerevisiae* region 1 and 4 showed small effects when inserted into the *S. uvarum* allele but large effects when replaced by the *S. uvarum* regions (Fig. 4-2, light blue vs. orange). It is possible that these two regions of *S. cerevisiae* contain large-effect copper-resistant changes that depend on the presence of other *S. cerevisiae* regions. Alternatively, the *S. cerevisiae* region 1 and 4 may only contain small-effect changes, and the sensitivity of the UCCC and CCCU constructs was caused by deleterious effects of the *S. uvarum* regions. Our data could not distinguish these two possibilities, although

the linear model suggested that synergistic epistasis between the *S. cerevisiae* region 1 and 4 could be the best explanation (Table 4-1).

### *Evolution of copper resistance*

The evolutionary history of *CUP2* might shed light on evolution of copper resistance in *S. cerevisiae*. The *CUP2* coding sequences did not exhibit signatures of positive selection according to site-specific dN/dS models (Scannell et al. 2011) or McDonald-Kreitman tests (Doniger et al. 2008). However, it showed significant heterogeneity in the dN/dS ratio across *Saccharomyces* lineages ( $p=0.00523$  compared to a model of fixed rates), with the *S. cerevisiae* lineage showing the highest ratio (0.562), suggesting variation in selection pressure across lineages (Scannell et al. 2011). The gain of copper resistance of *S. cerevisiae* has been associated with its adaptation to vineyard environment, where copper has been used as a fungicide (Mortimer 2000). While this trait is variable within *S. cerevisiae*, suggesting recent adaptation, most tested strains of *S. paradoxus* and *S. uvarum* are sensitive (Kvitek et al. 2008; Dashko et al. 2016). Therefore, *S. cerevisiae* might have acquired copper-resistant changes since its divergence from the two species, prior to adaptation of wine strains to the vineyard. This view is supported by the observation that the *S. cerevisiae* oak allele showed much higher copper resistance than the *S. uvarum* allele of *CUP2*. While variation of copper resistance within *S. cerevisiae* was largely attributed to copy number variation of *CUP1* and *CUP2* (Fogel and Welch 1982; S.L. Chang et al. 2013), the interspecific divergence may have a more complex genetic architecture. We showed that multiple changes in *CUP2* contribute to copper resistance in the present study, but the sum of their effects did not account for the total difference between *S. cerevisiae* and *S.*

*uvarum* (Chapter 3). Fully elucidation of the genetic basis of copper resistance would require further genetic analysis between *Saccharomyces* species.

# **Chapter 5: Comparative thermal tolerance of *S. cerevisiae* and *S. uvarum*.**

Part 3 of this chapter was done in collaboration with Ping Liu, Kim Lorenz and Justin C. Fay.

## Abstract

Divergence in growth temperatures between *S. cerevisiae* and *S. uvarum* has been well characterized, while survival at lethal heat shock temperatures is a related trait that may share genetic determinants with high temperature growth. In this chapter, I first describe my characterization of heat shock survival of the two species at near-lethal temperatures (40-50°C). I found that *S. uvarum* showed lower survival rates than *S. cerevisiae* during heat shock treatments, which mirrors its sensitivity to grow at high temperatures. Consistent with the divergence in heat shock survival, trehalose synthesis was induced in *S. uvarum* at a lower temperature than *S. cerevisiae*. While supplementing oleic acid slightly increased *S. uvarum*'s heat shock survival, the small effect suggests the heat sensitivity of *S. uvarum* cannot be simply explained by a difference in membrane lipid composition. Finally, we performed a genome-wide screen for *S. cerevisiae* genes that can enhance *S. uvarum*'s growth at high temperature using a molecular barcoded yeast (MoBY) ORF library. The screen was found to be confounded by an interaction between plasmid burden and temperature, but a screening strategy with repeated heat shock might potentially solve the problem. Taken together, the comparison between *S. cerevisiae* and *S. uvarum* in their heat shock survival and related physiological processes suggest that the two species diverged in signaling processes upstream of heat shock pathways and the divergence is likely to have a polygenic basis.

## **Part 1. Correlation between high-temperature growth and heat shock response.**

### **Introduction**

In the preceding chapters, my study on heat tolerance has been primarily focused on the yeasts' ability to *reproduce* at high temperature such as 37°C, often termed high-temperature growth (Htg) in the literature (Sinha et al. 2008). Heat tolerance can also be evaluated by the ability to *survive* exposure to a sudden lethal heat shock (HS), termed thermotolerance (Verghese et al. 2012). Most studies on thermotolerance have focused on acquired thermotolerance, which refers to a gain in survival at lethal temperatures after exposure to mild heat such as 37°C. Genes required for Htg, intrinsic thermotolerance (survival of lethal HS without prior exposure to mild heat) and acquired thermotolerance have limited overlaps according to a deletion screen (Jarolim et al. 2013), but there are shared molecular players among the three processes, particularly the genes involved in the heat shock response. For example, induction of heat shock proteins such as *SSA1*, *SSA2*, and *HSP104* can protect *S. cerevisiae* at lethal HS temperatures, while *SSA1* and *SSA2* in combination are also essential to 37°C growth (Craig and Jacobsen 1984). Disruption of genes in the cell wall integrity (CWI) pathway, Hsp90 cochaperones and heat shock transcription factor *HSF1* similarly causes sensitivity to 37°C (Verghese et al. 2012).

Despite these correlations in mutant studies, few studies have examined the relationship between Htg and HS survival in natural populations or different species. One study found that *S. cerevisiae* × *S. uvarum* F2 recombinants evolved under 31-46.5°C also showed increased survival at 48°C, demonstrating the two traits are likely linked; it was also the first study to

report *S. uvarum*'s hypersensitivity to 48°C (Piotrowski et al. 2012). A detailed characterization of the HS responses at lethal temperatures in *S. cerevisiae* and *S. uvarum* may provide insight in the physiology of thermal divergence and help identify candidate genes.

In addition to the expression of heat shock proteins, trehalose synthesis is another important process induced upon acute heat shock. Trehalose is a nonreducing disaccharide and was suggested to provide thermotolerance by stabilizing proteins and minimizing the effects of desiccation (Verghese et al. 2012), although a recent study provided a piece of contradicting evidence (Petitjean et al. 2015). Given the correlation between cellular trehalose level and thermotolerance, one possible explanation for *S. uvarum*'s thermosensitivity is a defect in the trehalose synthesis pathway. This hypothesis can be tested by quantification of trehalose levels of the two species at heat shock temperatures. Here, I report the characterization of the heat shock responses of *S. cerevisiae* and *S. uvarum* to lethal temperatures, including their survival rates at near-lethal temperatures (40°C to 50°C) and their trehalose content. I find that *S. uvarum* showed higher sensitivity to lethal heat shock temperatures, which mirrors its sensitivity in Htg, but the sensitivities could not be explained by a defect in trehalose synthesis.

## Materials and Methods

*S. cerevisiae* (YJF153, *MATa hoΔ::dsdAMX4*, YPS163 derivative) and *S. uvarum* (YJF1449, *MATa hoΔ::NatMX*) were grown to mid-log phase at room temperature in liquid YPD. The cultures were aliquoted to 2 ml per tube and shaken at designated heat shock temperatures at 250 rpm for 1 h. Every 15 min, one aliquot was taken out, diluted and plated onto YPD agar plates. OD readings were taken at the beginning and the end of the heat shock treatment. Viable cells were counted by colony formation at room temperature after 3-4 days. When there were

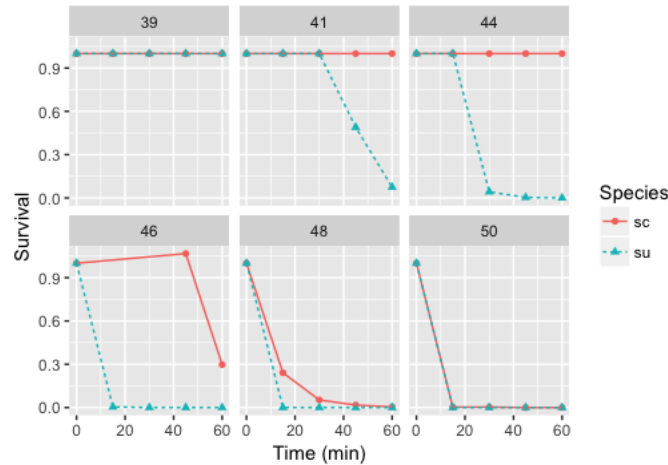


approximately the same number of colonies on the treatment plate as the control plate (0 min), colonies were not counted, and the survival rate was recorded as 1. For the 50°C treatment, colonies on the 0 min plate were not counted but estimated by OD readings (assuming 1 OD =  $5 \times 10^7$  cells/ml).

Trehalose level was measured at the end of the heat treatment: cells were harvested, washed twice in water and boiled in 200 ul water at 99°C for 10 min. The supernatant was analyzed by a trehalose assay kit (Megazyme) following manufacturer instructions. Trehalose was quantified by changes in 340 nm absorbance ( $\Delta A$ ) caused by NADPH production upon trehalose hydrolysis. Measurements that resulted in a  $\Delta A$  of less than 0.1 were retreated as no trehalose because they were under the detection limit. Trehalose levels were normalized to dry cell weight, calculated by: cell dry weight (g/L) =  $0.2 \times OD_{600}$  (Mahmud et al. 2010).

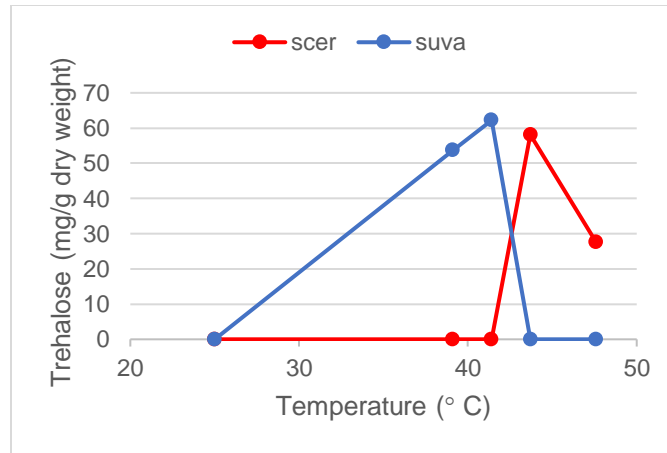
## Results

*S. cerevisiae* and *S. uvarum* was exposed to temperatures ranging from 39°C to 50°C and their survival rate during a 1-hour time course was measured by colony formation (Fig. 5-1). Consistent with the divergence in Htg, *S. uvarum* was more sensitive to heat shock than *S. cerevisiae*. 1 hour exposure to 41°C or above is lethal to *S. uvarum*, while 46°C or above is lethal to *S. cerevisiae*. 15 min at 46°C is sufficient to kill more than 99% of *S. uvarum* cells, but 50°C is needed to kill *S. cerevisiae* within 15 min. Therefore, *S. uvarum* not only cannot divide at high temperatures such as 37°C, but also cannot survive a short exposure to 41-46°C, temperatures at which *S. cerevisiae* is tolerant. It suggests the heat shock response diverged between *S. cerevisiae* and *S. uvarum* in parallel or via the same mechanism as their growth temperature divergence.



**Figure 5-1. Heat shock survival of *S. cerevisiae* (sc) and *S. uvarum* (su).** Numbers above each panel indicate the heat shock temperatures (°C). Survival is the ratio of alive cells after treatment to the total number of cells (time 0).

I further characterized the heat shock response of the two species by measuring trehalose production after 1-hr heat exposure. Consistent with their different survival rates, trehalose production was induced in *S. uvarum* and *S. cerevisiae* at different temperatures: *S. uvarum* induced trehalose production at 39-41°C, while *S. cerevisiae* did so at 44°C and above. *S. uvarum* did not produce trehalose at 44°C or above, presumably because the high temperatures had an immediate lethal effect on the species. Similarly, the trehalose production in *S. cerevisiae* at 48°C is lower than that at 44°C, reflecting an upper limit of heat shock response.



**Figure 5-2. Trehalose production of *S. cerevisiae* (scer) and *S. uvarum* (suva) at different temperatures (25°, 39°C, 41°C, 44°C and 48°C).**

## Discussion

In addition to the divergence in high temperature growth, *S. cerevisiae* and *S. uvarum* diverged in thermotolerance to acute heat shock, with the lethal temperature of *S. uvarum* being 4°C lower than *S. cerevisiae* (44°C vs. 48°C). The temperatures at which trehalose production peaked were also 4°C apart in the two species, suggesting that the divergence in thermotolerance occurred upstream of the trehalose synthesis pathway, likely in the sensing of heat stress. For example, there might be misfolded proteins accumulated in *S. uvarum* that triggered the heat shock response under mild heat conditions. Given *S. uvarum*'s cryotolerance, it is possible that its proteins are adapted to fold at cold temperatures and are unstable at high temperatures. Further comparison of misfolding or thermostabilities of proteins in the two species might help identify the contributors to thermal divergence. In addition, changes in the level of saturation in membrane lipids have also been proposed to trigger heat shock response (Carratù et al. 1996) and

could potentially contribute to the divergence in heat sensing. The role of lipid composition will thus be further examined in Part 2 of this Chapter.

## **Part 2. Effects of lipid supplements on thermotolerance.**

### **Introduction**

Temperature can significantly impact membrane fluidity: heat makes the membrane more fluid while cold makes it more rigid and viscous (Vigh et al. 1998; Beney and Gervais 2001). To counteract the temperature effect and keep their membrane fluidity at a normal level, organisms living in different thermal environment have evolved different lipid compositions: thermophilic yeasts have a high percentage of saturated fatty acid (SFA) while the mesophilic and psychrophilic yeasts have high levels of unsaturated fatty acid (UFA) (Arthur and Watson 1976). UFAs with unsaturated double bonds make the membrane more fluid and thus could mitigate cold-induced membrane rigidity, while SFAs could help with heat tolerance by stabilizing the membrane. Upon acute heat shock or cold shock, changes in membrane fluidity can also trigger transcriptional response, inducing either protein chaperones (heat shock) or fatty acid desaturase (cold shock) (Carratù et al. 1996; Vigh et al. 1998). Therefore, the lipid composition of cell membrane might be critical to thermal divergence in yeast.

Given the correlation between SFA content and thermotolerance, one may predict that fatty acids with fewer unsaturated double bonds could increase thermotolerance more. Indeed, supplementing the *S. cerevisiae ole1* mutant with oleic acid (C<sub>18:1</sub>) rendered higher thermotolerance than with linoleic (C<sub>18:2</sub>) and linolenic (C<sub>18:3</sub>) acids (subscript indicates the ratio of the number of carbon atoms to the number of unsaturated linkages) (Swan and Watson 1999).

Similar supplementation experiments could be performed with *S. uvarum*, to test whether its thermosensitivity can be rescued to a comparable level as that of *S. cerevisiae*. If so, it would indicate that a difference in membrane lipid composition may be a critical component of thermal divergence between *S. cerevisiae* and *S. uvarum*. In addition to SFA, ergosterol has also been suggested to stabilize membranes against heat and ethanol stresses (Swan and Watson 1998), and its effect on *S. uvarum* will also be tested in this section.

## Materials and Methods

Overnight cultures of two *S. cerevisiae* strains (YJF153, *MATa hoΔ::dsdAMX4*, YPS163 derivative; YJF641, *MATa ura3Δ leu2Δ his3Δ met15Δ*, BY4741 derivative) and one *S. uvarum* strain (YJF1449) were used to inoculate 5 ml CM with lipid supplements (0.25% or 0.5% Tween 80; 1, 10, or 20 ug/ml ergosterol) in glass tubes. Tween 80 contains 70% oleic acid. The ergosterol-supplemented medium was prepared from a 1mg/ml ergosterol stock solution, which contained 95% ethanol. The cultures were grown at 30°C for 4h to mid-log phase, then aliquoted and placed at 48°C (2 ml) or 30°C (2.5 ml) for 15 min. Cultures in CM without any supplements were diluted and plated onto YPD to examine heat shock effects. Thermotolerance was measured by the amount of growth after heat shock: the treated cultures were 1:10 diluted in fresh CM and grown overnight at room temperature at 1,200 rpm, with their OD<sub>600</sub> taken every 3 min by an iEMS plate reader. *grofit* package (Kahm et al. 2010) in R was used to analyze growth curves, and area under curve (AUC) was used to represent mean growth. Linear model was used to analyze the effects oleic acid or ergosterol concentration and heat shock (HS) on the strains:  $AUC \sim HS + concentration + HS*concentration$ , where \* indicates interaction.

## Results and Discussion

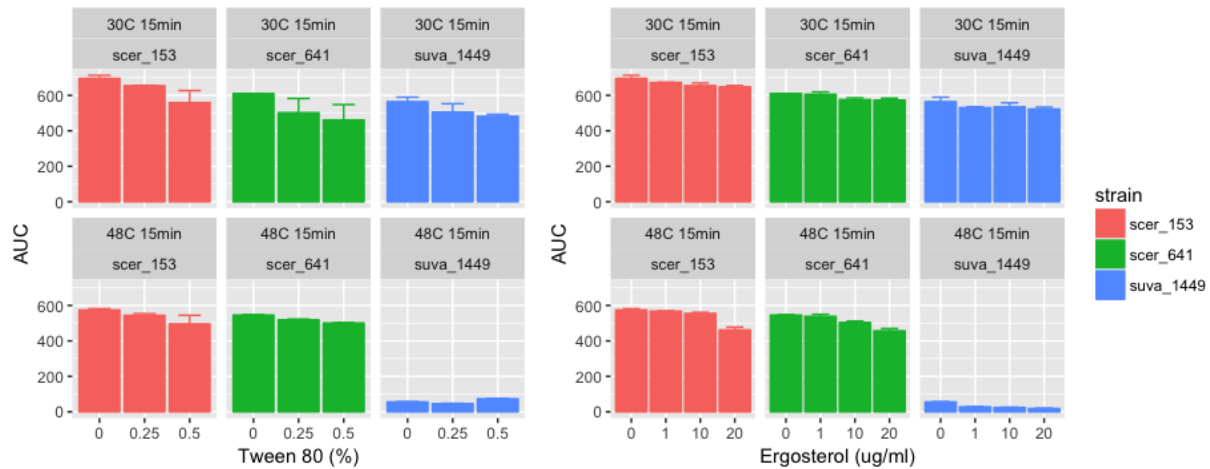
*S. cerevisiae* and *S. uvarum* supplemented with oleic acid and ergosterol was exposed to 48°C for 15min to measure their thermotolerance. This condition was shown to kill more than 70% of cells in a prior experiment (Part 1), but in this experiment, colony counts of non-supplemented cultures indicate approximately 100% of *S. cerevisiae* cells and 43% of *S. uvarum* cells survived, presumably due to differences in pre-treatment conditions (30°C instead of room temperature; different media). The heat shock effects measured from after-shock growth curves were consistent with the colony counts: the HS had a marginal effect on *S. cerevisiae* and a large effect on *S. uvarum* (Fig. 5-3). Therefore, this experiment only allowed me to evaluate the effects of lipid supplements on thermotolerance of *S. uvarum*.

Oleic acid had a general negative effect on growth ( $p < 0.05$ ), but increased *S. uvarum*'s survival after 15min's heat shock ( $p = 0.0433$  for *HS \* concentration*; Fig. 5-3). This provides some evidence that the thermosensitivity of *S. uvarum* is related to low levels of SFA. However, a control treatment with *S. cerevisiae* at its lethal temperature is needed for comparison. Also, the increased thermotolerance of *S. uvarum* remained incomparable to that of *S. cerevisiae*, suggesting *S. uvarum*'s thermosensitivity cannot be simply explained by a difference in membrane lipid composition.

Ergosterol similarly had a negative growth effect on the two *S. cerevisiae* strains ( $p < 0.05$ ). The effect was especially obvious for the 20 ug/ml-48°C treatment (Fig. 5-3). One potential explanation is that ethanol in the ergosterol stock solution caused additional stress, although the negative effect was present even when the data in the 20ug/ml group was excluded.

In terms of its effect on thermotolerance, ergosterol did not increase heat shock survival for either *S. cerevisiae* or *S. uvarum*.

One limitation of the current experimental design is that the actual lipid composition of *S. uvarum* cells was not measured. In both of the previous reports, *S. cerevisiae* lipid auxotrophs (*ole1* and *erg1*) were used in order to maximize the intake of oleic acid or ergosterol, and the lipid composition of treated cells was determined by biochemical methods (Swan and Watson 1998; Swan and Watson 1999). The use of auxotrophs could be important, because the *ole1* strain was found to be highly amenable to changes in membrane lipid composition in a consistent and predictable manner (Swan and Watson 1999). In the current study, it was unknown how much of the supplemented lipids was incorporated into *S. uvarum*, making it hard to evaluate their effects.



**Figure 5-3. Effects of oleic acid (left) and ergosterol (right) on thermotolerance of *S. cerevisiae* (scer\_153 and scer\_641) and *S. uvarum* (suva\_1449).** Thermotolerance was represented by the amount of growth after heat shock (AUC, area under the curve). Error bars represent the standard error of two replicates.

## **Part 3. Screen for *S. cerevisiae* genes that enhance *S. uvarum*'s heat tolerance in *S. uvarum* using the MoBY-ORF library.**

### **Introduction**

Dissecting the genetic basis of species' differences has been challenging, largely due to reproductive barriers between species. Traditional quantitative genetics approach such as QTL mapping can only be applied to closely related species and it is often limited in statistical power and resolution. Candidate approaches have provided important insight into the molecular mechanisms of evolution, but they often focus on single genes of large effect and do not address questions such as the number of genes involved and the distribution of effect sizes. With the abundant resources and ease of genetic manipulation, yeast offers the ability to design high-throughput screens to systematically dissect the genetic basis of interspecific differences. In Chapter 3, I described a non-complementation screen for genes underlying thermal divergence between *S. cerevisiae* and *S. uvarum*. However, the non-complementation screen does not cover essential genes and cannot identify genes whose effects are masked by other genes in the *S. cerevisiae* genome, i.e. redundant heat-tolerant alleles. An alternative approach is to screen for *S. cerevisiae* genes that enhance heat tolerance in *S. uvarum*. This strategy could cover essential genes and identify any single genes with a detectable additive effect regardless of redundancy, thus complementing the first screen.

We use a previously constructed molecular barcoded yeast ORF library (Ho et al. 2009), hereafter referred to as the MoBY library to screen for *S. cerevisiae* alleles conferring heat tolerance. The MoBY library is a collection of 4,956 centromere-based plasmids, and each plasmid carries one *S. cerevisiae* gene along with two unique oligonucleotide barcodes. The



screening strategy can be combined with next-generation sequencing techniques to achieve higher throughput and resolution. By pooled transformation of the MoBY library and barcode sequencing, we were able to measure the fitness effects of thousands of *S. cerevisiae* genes in *S. uvarum* under various conditions. However, during validation of the candidate genes coming out of the screen, we found that the MoBY plasmids have plasmid burden in *S. uvarum*, that plasmid burden increases with temperature, and that plasmids likely differ in their burden depending on the inserted sequence, thus confound the screening results. A heat-shock based strategy that is independent of cell replication could be considered as an alternative screening approach.

## **Materials and Methods**

### *Transformation of MoBY-ORF library and pooling*

The MoBY-ORF library was grown following (Ho et al. 2009), pooled and then minipreped. The pooled plasmids were transformed into competent cells of *S. uvarum* (YJF1449, *MATa hoΔ::NatMX*), prepared according to (Dohmen et al. 1991). The transformation followed the lithium acetate protocol (Gietz et al. 1995) with modified heat shock (37°C) and incubation (23°C) temperatures. Transformants were selected on YPD+G418 plates; a total of 62,000 colonies were pooled for growth competition.

### *Growth competition*

The pooled transformants were competed under twelve treatments: 23°C, 30°C, 33°C and each of YPD+G418 with 0, 4, 5 or 6% ethanol. Following overnight culture in YPD+G418 at 30°C, the pooled transformants were inoculated into each treatment condition (OD = 0.1) with three replicates per treatment. Each culture was diluted to OD of 0.1 in fresh medium after 24

and 48 hours; if the OD did not reach 0.1, the culture was resuspended in fresh medium instead. Samples were taken at 0, 24, 48 and 72 hours for barcode-sequencing. Little growth occurred for the cultures in 6% ethanol at 30°C and all cultures at 33°C; the former was removed from further analysis and the latter was only examined for the 72h data. For the remaining samples, OD600 measurements indicated an average of 1.7 to 6.7 doublings per day. A total of 78 samples were subjected to sequencing: 3 (t=0) + 3 timepoints \* 2 temperatures \* 4 ethanols \* 3 replicates – 9 samples of 30°C at 6% ethanol + 12 33°C samples which were only collected at 72h.

### *Barcoded sequencing and data analysis*

A combination of twenty 8-mer barcodes and 4 index primers were used for pooled sequencing of all the samples. Each sample was designed to have a unique barcode-index combination. The MoBY uptag andowntag were amplified from each sample at 25 cycles (30 cycles for 15 samples with low DNA concentration), using one barcoded primer targeting the tag and one primer targeting *KanMX*. Both primers contained homology to Illumina indexing primers. The barcoded PCR products were pooled, and the Illumina indexes were added by a second PCR (20 cycles). After indexing, the pools were combined at equal concentrations and sequenced on a single Illumina HiSeq lane, yielding a total of 222 million reads.

We accepted up to 2 mismatches in the index (7 bp length) and 1 mismatch in the barcode (8 bp) when demultiplexing the data. Constant primer regions adjacent to up/down tags were matched. When matched, we extracted up/down tags and accepted hits if they were up to 7 mismatches (out of 20) from a single MoBY tag. If there were multiple hits for a given number of mismatches, we did not count the read. There were 4938 uptags and 4892owntags that were usable, a total of 4981 genes. A total of 119 million tags across 78 libraries were usable.

DESeq was used to center and normalize the log transformed data using the 'pooled' variance stabilization. Uptags and downtag counts were combined. 564 genes were removed because of an average of 10 or fewer reads per sample. 622 genes were removed because of 20 or more samples (out of 78) with no reads. These treatments left a total of 4317 genes for analysis and an average of 1.5 million reads per sample (the lowest being 480k) and a median of 145 reads per gene.

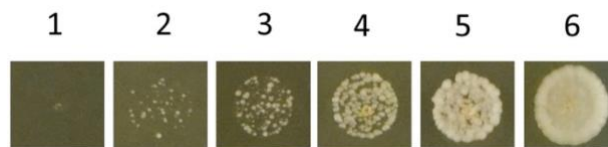
The 23°C and 30°C data were fit to two linear models using R: (1)  $\text{data} \sim \text{temp} + \text{media} + \text{time} + \text{temp}:\text{time} + \text{media}:\text{time} + \text{temp}:\text{media}:\text{time}$ ; (2)  $\text{data} \sim \text{temp} + \text{media} + \text{temp}:\text{media}:\text{time}$ , where a colon indicates an interaction; raw log counts were used as weights in the regression. The second model treats time as a factor rather than numeric, which tested for effects for each combination of temperature and medium. This model doesn't assume a linear relationship of the data over time, e.g. it can go up then down. As a control, we permuted the labels and retested the data. For each of three permutations we identified fewer than 9 false positives for each test at a p-value cutoff of 0.001, as expected for 4317 tests.

The 33°C data only had one timepoint (72h) so we analyzed it separately. We applied the same cutoffs of 9 or fewer average counts and at least 25% of samples having no zero data, leaving 4320 genes in the dataset. We applied two models for all of the 72h data: (1)  $\text{data} \sim \text{temp} + \text{media} + \text{media}:\text{temp}$ ; (2)  $\text{data} \sim \text{temp} * \text{media}$ , where the second model treats media and temp as factors.

### *Validation of plasmid phenotype*

We examined genes that had significant temperature effects in three models: 1) 22°C and 30°C data, taking temperature as a numeric variable; 2) 22°C and 30°C data, taking temperature

as a factorial variable; 3) 22°C, 30°C and 33°C growth at 72hr, taking temperature as a factor. There were 128 such genes. Using Sanger sequencing we confirmed 80 of 96 selected plasmids to be correct. 78 of the correct plasmids were individually transformed into a haploid *S. uvarum* strain (YJF1449). The transformants were spotted onto YPD+G418 agar plates and their growth at 33.5°C after 4 days was manually scored by three individuals using a universal reference (Fig. 5-4). The spot assay was repeated four times in two separate days. Pictures of the four replicates were scored together and the inter-person variability was small (<0.75). The resulting 12 scores (3 people × 4 replicates) were averaged to represent a strain's heat tolerance.



**Figure 5-4. Scoring scheme for MoBY transformants.**

#### *Allele substitution of candidate genes*

41 *S. cerevisiae* genes enhanced *S. uvarum*'s heat tolerance on plasmid, among which we selected a subset to validate individually by allele substitution. Our major strategy was to knock out the *S. uvarum* allele at its endogenous locus and integrate the *S. cerevisiae* allele at *ura3* locus. We first generated an *S. uvarum* strain that is compatible with the integrative plasmid pRS306, by replacing *S. uvarum* *URA3* with *S. cerevisiae* *ura3-140* (*Scer-ura3-140*) allele (Engle and Fay 2013). The *Scer-ura3-140* allele was amplified from YJF149, including 299 bp upstream and 186 bp downstream intergenic sequences. Homologous arms to *S. uvarum* *URA3* intergenic regions were added by a second PCR (forward primer 5'-

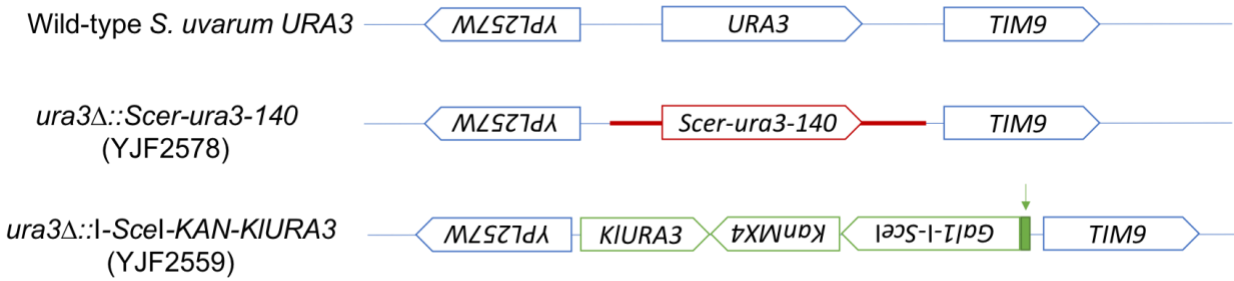
TGGTGTTGTTCTTATGTAACTATATGTAATGAGCGCCCCTTTCTGCGAGGCATATTT

ATGGTGAAGG-3', reverse primer 5'-

GTTTGGATGTCTTTAAATTTGCTATGATATTTGGATAATTCCTTGGTTCTGGCGAG

GTATTGG-3'; underlined bases anneal to the first-round PCR product). The PCR product was transformed into *S. uvarum* to create the pRS306-compatible strain YJF2578. Both coding and flanking sequences of *S. uvarum URA3* were replaced in this strain: only 80 and 86 bp *S. uvarum* intergenic sequences that are immediately adjacent to the upstream and downstream ORFs (*YPL257W* and *TIM9*), respectively, were left (Fig. 5-5). The resulting *ura3Δ::Scer-ura3-140* locus was confirmed by Sanger sequencing (Appendix).

To increase transformation efficiency and streamline the experimental process, we further replaced the *ura3-140* allele with the CORE-I-*SceI* cassette (*Gall-I-SceI kanMX4 KlURA3*) from pGSKU (Storici et al. 2003) to create *S. uvarum* strain YJF2559 (forward 5'-  
ttagtgcgctccatctctgcggtcctgtcttgttctgtttggtctgagccTTCGTACGCTGCAGGTCGAC-3', reverse 5'-  
tcttgcgctcacaccaccattgttctccagtaggggtactatagaataaa**TAGGGATAACAGGGTAATCCGCGCGT**  
TGGCCGATTCAT-3'; *S. uvarum* sequences are in lower case, I-*SceI* recognition site is in bold and P.I and P.IIS primers provided by (Stuckey et al. 2011) are in upper case). After the replacement, 30 and 36 bp intergenic sequences flank the CORE-I-*SceI* cassette at its 5' and 3' end, respectively (Fig. 5-5). This design raised concern at a later timepoint because there was only 36 bp left of the endogenous promoter of *TIM9*, the function of which is essential to mitochondrial function. However, we tested and found YPEG growth of YJF2559 was not affected. YJF2559 did exhibit increased heat sensitivity compared to YJF1449 and YJF2578, but YJF2561, another transformant in this experiment, was not affected, suggesting cryptic genetic changes in YJF2559, rather than the construct design, conferred heat sensitivity.



**Figure 5-5. Diagram of *ura3* alleles constructed in this study.** Blue, *S. uvarum* sequences; red, *S. cerevisiae* sequences; green, pCORE sequences. The green arrow indicates the I-SceI recognition site. Boxes represent ORFs while lines represent non-coding sequences. Note that *URA3* and *TIM9* are on the reverse strand and *YPL257W* is on the forward strand in the *S. uvarum* reference genome (Scannell et al. 2011), but this diagram shows *URA3* as on the sense strand for the ease of visualization. *YPL257W* is referred to as the upstream gene and *TIM9* as the downstream gene of *URA3*.

8 candidate genes (*INP51*, *DSE1*, *SPO71*, *APM1*, *AZF1*, *MSS51*, *YJRO98C*, and *GND2*) were successfully knocked out with *hygMX* in *S. uvarum* (YJF2559, *MATa ho::NatMX ura3Δ::I-SceI-KAN-KIURA3*), using orthologs defined by Scannell et al. (2011). The *S. cerevisiae* alleles for replacement were designed to match the regions covered by the MoBY plasmids, with modifications ranging from +124 bp to -16 bp at ends for primer design concerns. The *S. cerevisiae* alleles were PCR-amplified off genomic DNA (YJF153) using primers with 20bp-homology to sequences flanking the CORE-I-SceI cassette at *S. uvarum URA3* locus, with Q5 polymerase (New England Biolabs). The homology was extended to ~45 bp with a second PCR. In the cases of difficult PCRs, the *S. cerevisiae* alleles were amplified in two pieces with ~200 bp overlap and transformed together into *S. uvarum* for gap repair. The PCR products were

transformed into corresponding *S. uvarum* knockout strains following the galactose-induction protocol: overnight cultures of *S. uvarum* strains were harvested, washed, resuspended in YP-Raffinose and grown at 20°C at 250 rpm for 4h. Galactose was then added to the cultures at a final concentration of 2%, and the cells were shaken for 6.5h to induce double-strand break at the I-SceI site. The cultures were then harvested and transformed with the lithium acetate method. Successful replacement of the CORE-I-SceI cassette (containing *KIURA3* marker) was selected by 5-FOA. Using this strategy, we obtained allele substitutions for 4 candidate genes (*MSS51*, *YJR098C*, *GND2*, *SPO71*). *S. uvarum* alleles were integrated in parallel as controls, with primers that were designed based on alignment to the amplified *S. cerevisiae* alleles. Integration of the *S. cerevisiae* alleles were confirmed by PCR at both junctions, while integration of the *S. uvarum* alleles were confirmed at one junction (*SPO71*, *YJR098C*, *GND2*) or not confirmed by PCR (*MSS51*).

We also generated an allele substitution for *IPT1*, a positive hit in the pooled competition but not validated by the individual plasmid assay. We knocked out *IPT1* with *hygMX* in *S. uvarum* (YJF2578). The *S. cerevisiae* (YJF153) allele, including the coding sequence, the entire 5' and 79 bp of 3' intergenic sequences, was Gibson-assembled into pRS306, which was linearized by *StuI* and integrated at the *ura3* locus.

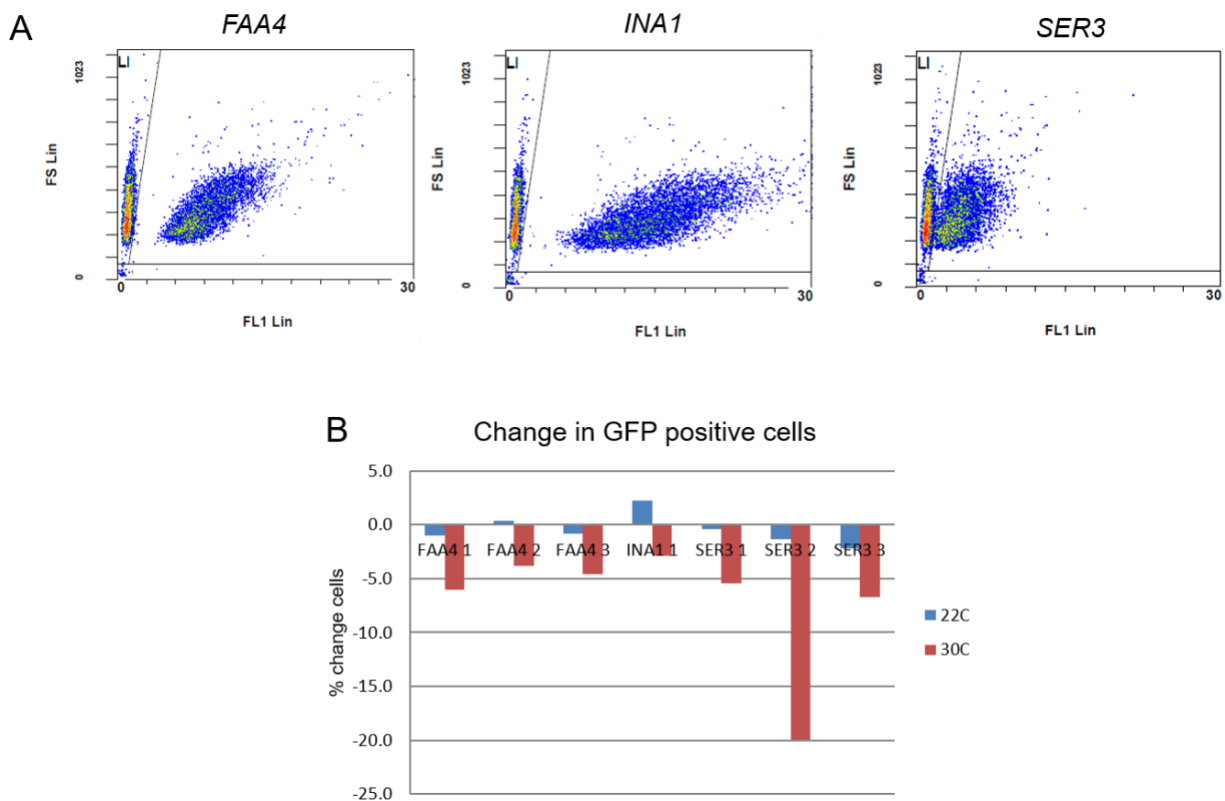
For two of the top candidates (*NPL6* and *SMY2*), we knocked out the *S. uvarum* alleles with the CORE-I-SceI cassette (*Gall-I-SceI hyg KIURA3*) from pGSHU (Storici et al. 2003) in YJF2578 (*MATa hoΔ::NatMX ura3Δ::Scer-ura3-140*). The replacement knocked out both the coding and the flanking intergenic sequences. We then replaced the CORE-I-SceI cassette with the *S. cerevisiae* alleles, including both coding and intergenic sequences from YJF153, using the

galactose-induction protocol described above. To generate homozygous diploids for growth competition, the transformants were mated to an *S. uvarum* strain (YJF2539, *MAT $\alpha$*  *ho $\Delta$ ::KanMX*). The diploids were sporulated and haploids carrying the *S. cerevisiae* allele were mated to each other to construct diploid *S. uvarum* strains that were homozygous for *S. cerevisiae* alleles at the candidate locus.

#### *GFP strain construction and growth competition*

Diploid strains of *S. uvarum* were required for flow cytometry because they are not flocculent. GFP-*KanMX4* cassette was PCR-amplified from pKT127 and transformed into diploid *S. uvarum* (YJF2602) to replace the stop codon of *FAA1*, *INA1* or *SER3*, genes selected based on *S. cerevisiae* literature. The GFP-tagged strains were competed with the parental strain (YJF2602) to test their ability to separate GFP positive vs. negative cells, as well as their fitness, using a Beckman Coulter FC 500 MPL flow cytometer (Beckman Coulter, Brea, CA). The *FAA4*-fused GFP showed the best separation (Fig. 5-6A) and only slight fitness defect at 30°C (Fig. 5-6B), thus used in subsequent fitness assays.





**Figure 5-6. GFP-tagged *S. uvarum* strains and their competition with the parental strain.**

**A.** *S. uvarum* carrying GFP-tagged *FAA4*, *INA1* or *SER3* was mixed with GFP-less strain YJF2602 and examined by a flow cytometer. The *FAA4*-GFP fusion showed the best separation of GFP positive vs. negative cells. **B.** Change in the percentage of GFP positive cells after 24 hr growth at designated temperatures. 3 replicates for the *FAA4* and *SER3* marker strains and 1 replicate for the *INA1* strain are shown.

The *FAA4*-GFP fusion strain (YJF2656) was competed with strains with replaced alleles to determine the latter's fitness. Overnight cultures of the GFP and non-GFP strains were 1:1 mixed based on ODs. The mixtures were used to inoculate 500 ul fresh YPD at a starting OD of 0.02. The strains were competed at 1) 22°C for 23h, 2) 33°C for 48h and 3) 33°C for 48h

followed by a 24h-recovery at 22°C, in deep well plates shaken at 400 rpm. In the third treatment, the cultures were diluted to OD of 0.01 in fresh YPD before the 24h-recovery. The cultures were sampled every 24h and assayed by flow cytometer to determine the ratio of GFP positive cells. Monocultures of the GFP strain were grown under the same conditions; the GFP false negative rate was 4-5% at 22°C and ~20% at 33°C. The counts of GFP-tagged cells ( $G$ ) and the non-GFP competitors ( $N$ ) were thus corrected using temperature-specific false negative rates ( $r$ ): the number of GFP cells that appeared GFP-negative was  $\Delta G = G_{measured} / (1 - r) - G_{measured}$ , where  $G_{measured}$  was the flow cytometer count. Then  $G$  and  $N$  were corrected by:  $G = G_{measured} + \Delta G$ , and  $N = N_{measured} - \Delta G$ .

Fitness was measured according to (Hartl et al. 1997). The growth difference between GFP negative and positive cells in a given time was  $w = \ln(n_1 / g_1) - \ln(n_0 / g_0)$ , where  $g_0$  and  $n_0$  are the corrected starting frequencies of GFP positive and negative cells, respectively, and  $g_1$  and  $n_1$  are the corrected frequencies at the end of the competition. Note that  $n_0 = 1 - g_0$ ,  $n_1 = 1 - g_1$ .

### *Integration of MoBY plasmids*

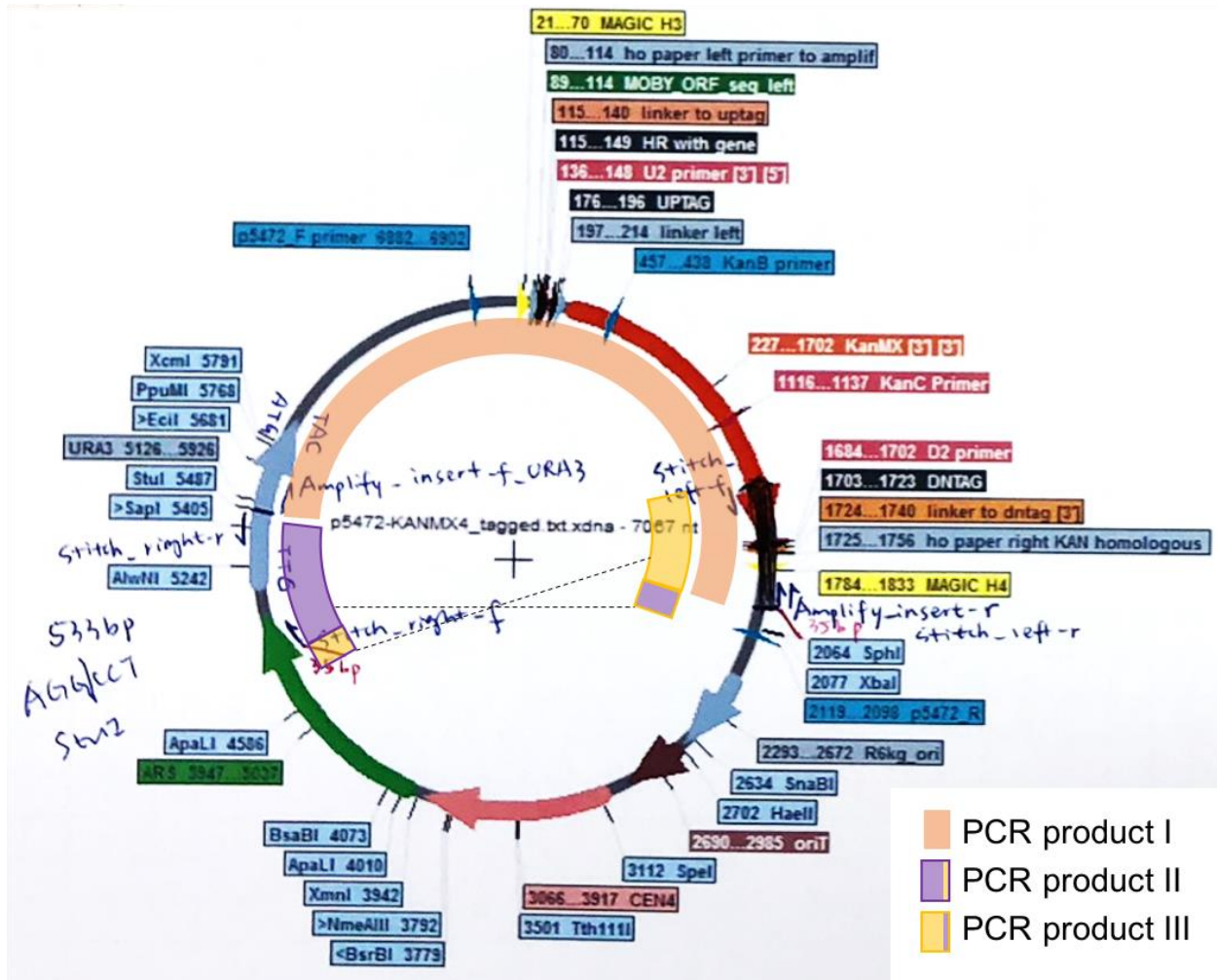
To examine the phenotypic effects of the candidate genes in a relatively high-throughput manner, I developed a PCR-based method to integrate MoBY plasmids into the *ura3* locus of *S. uvarum*. The integration required 1) removal of CEN sequence from the MoBY plasmid and 2) linearization of the plasmid by a cut in *URA3* sequence to allow homologous recombination with genomic *ura3*. The former can be achieved by gap repair and the latter can be achieved either by restriction digest or by PCR. Because the *S. cerevisiae* gene inserts in MoBY plasmids were diverse sequences, it was difficult to find a universal restriction enzyme to cut the plasmids. Also, restriction digest followed by gel-extraction required a large amount of plasmid DNA as

input. Therefore, I used stitching PCR and gap repair to assemble a linearized CEN-less MoBY plasmid, with halves of *URA3* gene at the ends of the linear molecule (Fig. 5-7).

Gap repair is a common yeast cloning method; two pieces of DNA with sufficient amount of homology can be assembled into one via yeast homologous recombination. I first PCR-amplified a segment of MoBY plasmid containing the 5' half of *URA3* and the *S. cerevisiae* gene insert (PCR product I, the orange band in Fig. 5-7). Here I used a minimum amount of minipreps (0.2-0.9 ng/ul) as template to avoid carry-over of CEN plasmids. The second piece of DNA should include the 3' half of *URA3* and some homology to the first piece of DNA, but “skip” the CEN sequence on MoBY plasmid. I achieved this by stitching PCR: the “left” piece for stitching (PCR product III, the yellow band in Fig. 5-7) is 306 bp, including sequence at the *KanMX* end of PCR product I and 35 bp homology to the “right” piece; the right piece (PCR product II, the purple band in Fig. 5-8) is 533 bp, including the 3' half of *URA3*, its adjacent sequence and 35 bp homology to the left piece. Both PCR product II and III were amplified from a MoBY plasmid without insert (0.3 ng/ul). Stitching was performed using Pfu Ultra II in two steps: 1) Purified PCR product II and III were mixed (9.75ul each) with polymerase, dNTPs and buffer solution in a 25 ul reaction and underwent 10 PCR cycles (95°C 2min; 95°C 20s, 52°C 30s, 72°C 55s for 10 cycles; 72°C 3min). 2) 20 ul of the step-one PCR reaction was mixed with primers (stitch\_left\_f, stitch\_right\_r, Table 5-1) and other PCR reagents to make a 50 ul PCR reaction; the reaction was run for 30 cycles with two annealing temperatures: 95°C 2min; 95°C 20s, 50°C 20s for 5 cycles + 60°C 20s for 25 cycles, 72°C 55s; 72°C 3min. The final product was run on an agarose gel and a band with expected size of 839 bp was cut and extracted, yielding PCR product IV used for gap repair. To get more DNA for ~30 transformations, the PCR product IV were further amplified using stitch\_left\_f and stitch\_right\_r.

For each gene, 50 ul PCR product I and 24 ul PCR product IV were transformed together into *S. uvarum* carrying *S. cerevisiae ura3-140* allele (YJF2578). The two DNA pieces were expected to recombine with each other via the ~300 bp homology (gap repair), as well as with *S. cerevisiae ura3-140*, and only repaired, integrated plasmid with intact *URA3* can be selected on CM-ura plates. The transformants were further tested for G418 resistance since the integrated plasmids contain *KanMX* sequence. In my initial test, 15 out of 21 Ura+ transformants were G418-resistant, and integration of the 15 transformants were all confirmed by PCR for at least one junction, indicating that G418 resistance was a reliable marker for successful integration. I then used this method to integrate 30 candidate plasmids into *S. uvarum*, along with one control plasmid without any insert. The candidates were from pooled competition results, and 10 of them showed positive effects in individual plasmid assays.

Three independent transformants were picked for each plasmid. High variation in 34°C growth among transformants was observed for some genotypes, and growth on YPEG plates indicated none of them were petites. To remove effects of any secondary mutations introduced during transformation, the transformants were mated to two *MAT $\alpha$*  *S. uvarum* strains (YJF2546 and YJF2547; *MAT $\alpha$  ho $\Delta$ ::hygMX*) and the diploids were selected on YPD+G418+Hyg plates. The haploids and diploids were pinned to YPD agar plates using a Singer ROTOR (Singer Instruments, Watchet UK) and assayed for growth at room temperature, 34°C and 35°C. ImageJ was used to measure the colony sizes.



**Figure 5-7. Diagram of MoBY plasmid.** The outer circle shows the MoBY plasmid backbone and inner circle shows the PCR fragments. The *S. cerevisiae* gene is inserted between the MAGIC H3 site (yellow) and *KanMX* (red) (insert is not shown). The first PCR (orange) covers the 5' half of *URA3* (blue arrow on the left), the *S. cerevisiae* insert, and *KanMX* (red). The second PCR (purple) covers the 3' half of *URA3*, with added homology to the third PCR product (yellow), which covers ~300bp of PCR product I. Dashed lines indicate assembly via stitching PCR. Note that the "left" piece of stitching PCR (PCR product III) is on the right and the "right" piece (PCR product II) is on the left in this diagram.

Table 5-1 Primers used in MoBY integration.

PCR	Primer	Sequence (5'->3') <sup>1</sup>
<b>I</b>	Amplify_insert-f_URA3	CCTCTAGGTTTCCTTTGTTACTTCT
<b>II</b>	stitch_PCR_right-f	tctattaccctgttatccctagagctcgctgatcaTGGCAAACCGA GGAACTCTT
<b>II, stitching (IV)</b>	stitch_PCR_right-r	CCTTTTGATGTTAGCAGAATTGTCA
<b>III, stitching (IV)</b>	stitch_PCR_left-f	GCCTCGACATCATCTGCCCA
<b>I, III</b>	Amplify_insert-r /stitch_PCR_left-r	cgtggcaagaataccaagagttcctcggtttgccaTGATCAGCG AGCTCTAGGGA

<sup>1</sup>Colored bases indicate added homology, following the same color scheme as in Fig. 5-7.

#### *Plasmid burden assay*

YJF2602-derived MoBY transformants were inoculated into liquid YPD and grown at 22°C or 30°C for 24h. The cultures were diluted and plated onto YPD and YPD+G418 plates at the beginning and the end of the growth. The number of generations during the 24h growth was calculated by  $n = \log ([I] / [F]) / \log (2)$ , where  $[I]$  and  $[F]$  are the initial and final cell concentration (number of cells per ml), respectively, calculated based on colony counts. The plasmid loss rate per generation was calculated as  $1 - \left( \frac{ratio_{End}}{ratio_{Start}} \right)^{-n}$ , where *ratio* is the number of colonies on G418 plates divided by that on YPD plates.

### *Heat shock-based selection on MoBY transformants*

Pooled MoBY transformants were recovered from -80°C by growing in liquid YPD+G418 at 22°C for 1 day. The culture was then inoculated into fresh YPD+G418 (OD=1) and grown for 2 days to reach stationary phase (OD=20). 100ul culture was added to 2ml pre-warmed YP+G418 liquid media in glass tubes and shaken at 250 rpm, 43°C for 1h. This culture is hereafter referred to as “HS group” and was always under heat shock (HS) treatment in each selection cycle. For the control treatment, the culture was inoculated into room temperature YP+G418 and left on bench for 1h (“mock group”). The same HS and mock treatment was performed with a diploid transformant of the MoBY vector as control. Note that YP was used during the treatment instead of YPD to repress growth. After the treatment, 3ml YPD+G418 was added to each tube and the cultures were grown at 22°C for 2-3 days to stationary phase again. This process was counted as the 1<sup>st</sup> selection cycle. The procedure was repeated for another four times, with a modified HS temperature of 44°C. Starting from the 3<sup>rd</sup> cycle, the mock treatment was changed to shaking at 250 rpm at 22°C for 1h. Starting from the 4<sup>th</sup> cycle, cultures in the mock group were 1:100 diluted in 2ml YP+G418 after the 1h treatment, to control for the number of generations and to mimic the bottleneck effect caused by heat shock.

At the 2<sup>nd</sup> and the 5<sup>th</sup> cycle, all samples were examined for their 44°C survival. For both HS and mock group, two cultures were set up for each sample and were placed at either 22°C or 44°C for 1h. (However, only the 22°C cultures of the mock group and the 44°C cultures of the HS group were grown to the next cycle.) At the end of the 1h treatment, the cultures were plated onto YPD and YPD+G418 agar plates. For each sample, the survival rate was calculated as the number of colonies grown after the 44°C treatment divided by that grown from the 22°C

treatment. Survival rates based on colony counts on YPD and YPD+G418 were similar (e.g. 1.18% vs 1.13%, 0.30% vs. 0.47% for the 2<sup>nd</sup> cycle), suggesting that HS affected the plasmid-bearing and plasmid-free cells the same. I thus used the counts from YPD+G418 when reporting the results.

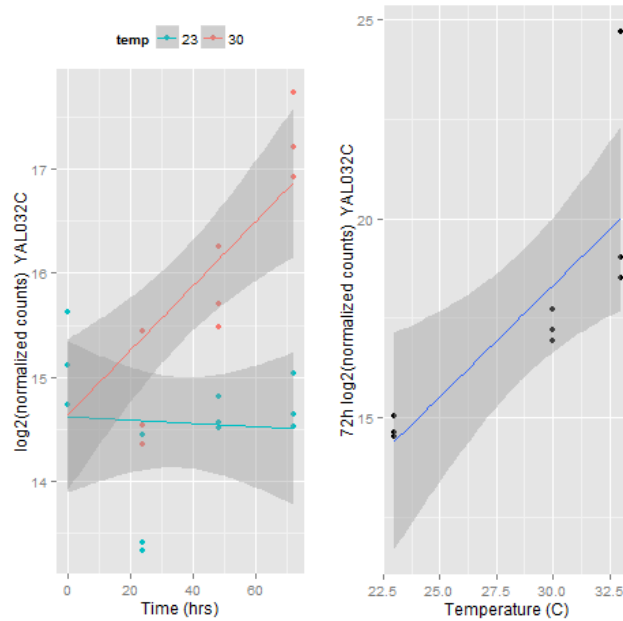
After the 5<sup>th</sup> selection cycle, the cultures were frozen as glycerol stocks at -80°C. As an assessment of the approach, 5 colonies were streaked from the heat-shocked transformant pool. The KanMX cassettes with flanking tags were PCR-amplified from the colonies and were Sanger-sequenced for the uptags. Survival after 1h-exposure to 44°C was measured for the 5 isolated colonies and a random colony from the mock pool, using the same condition as in the selection cycles. Additionally, to evaluate the possible effects of plasmid burden, HS survival of diploid MoBY transformants of *UME1*, *MCO76* and the vector control was measured.

## Results

### *Screen for S. cerevisiae genes that enhance S. uvarum's heat tolerance by growth competition*

We screened a pool of *S. uvarum* transformants carrying barcoded *S. cerevisiae* genes for increased heat tolerance by growth competition. The transformants were grown in YPD+G418 at 23°C, 30°C and 33°C for 72h, in combination with varied ethanol stresses (see Methods). The pool was sampled every 24h and were sequenced for their barcodes. We then used linear models to test for effects of time and temperature on the barcode abundance. Out of the 4317 genes detected in the pool, 84 genes showed increased frequency across time at 30°C compared to 23°C, as well as a positive temperature effect at 72h (see Fig. 5-8 for an example).

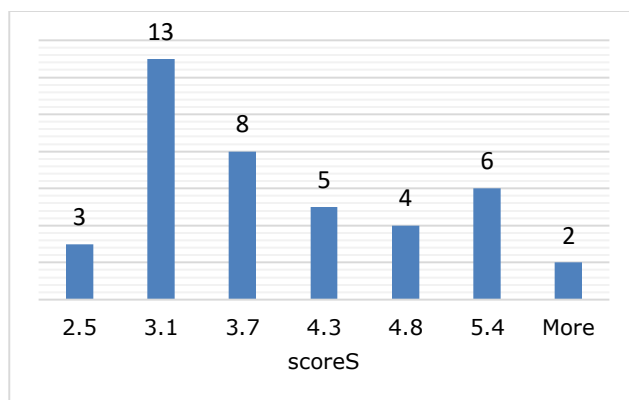




**Figure 5-8. A positive hit (*YAL032C/PRP45*) in the growth competition.** Abundance of the gene increased across time specifically at 30°C (left). It also showed a positive temperature effect at 72h (right).

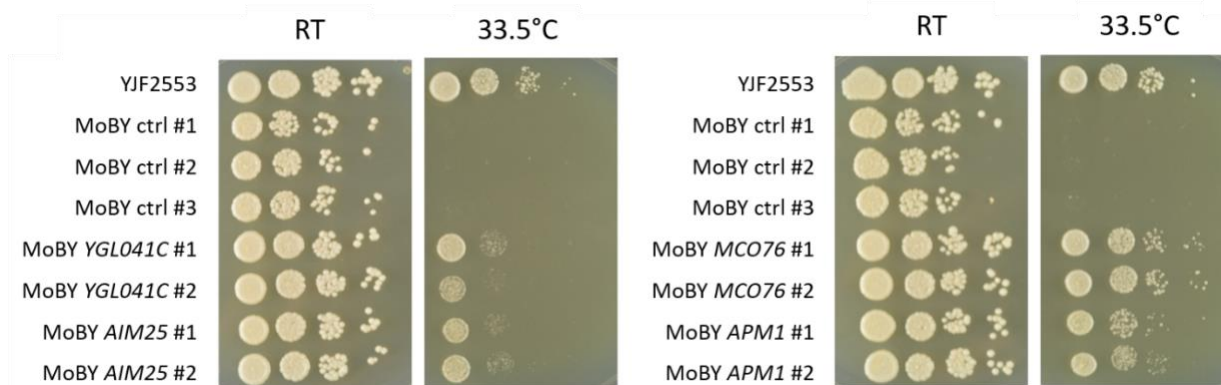
#### *Validation by individual plasmid transformation*

128 genes showed positive temperature effects across three linear models (see Methods) and were thus selected for individual validation. 78 sequencing-confirmed plasmids were transformed into haploid *S. uvarum* and tested for their ability to increase heat tolerance on YPD+G418 medium at 33.5°C by spot assays. Compared to the vector control, 41 genes showed higher growth scores (>2.5) at high temperature, mostly with small-to-moderate effects (Fig. 5-9). The top candidates include *NPL6*, *YPL109C*, *SMY2*, *YGL041C-B*, *GWT1*, and *INP51*, among which *NPL6* and *GWT1* have known heat sensitive phenotypes upon deletion.



**Figure 5-9. Distribution of growth scores of individual MoBY transformants that showed increased heat tolerance compared to the control plasmid.**

However, this validation had two caveats. When selecting the candidate gene set, we did not exclude genes that showed increased frequency across time at 23°C and we might have selected plasmids that conferred general growth benefits rather than temperature-specific effects. Also, all phenotypic assays were compared to the transformants of an empty MoBY vector, without a plasmid-free *S. uvarum* strain as control. Growth on G418 relies on plasmid maintenance, which may not be constant across transformants. We thus tested the heat tolerance of diploid *S. uvarum* transformed with candidate MoBY plasmids, in comparison to a G418+ *S. uvarum* diploid. Although the candidate plasmids (*YGL041C*, *AIM25*, *MCO76*, *APM1*) conferred heat tolerance compared to the vector control, none of them grew significantly better than the *S. uvarum* diploid with *KanMX* integrated (Fig. 5-10), implying that the *S. cerevisiae* inserts on these plasmids might have mitigated the plasmid burden rather than the heat sensitivity.

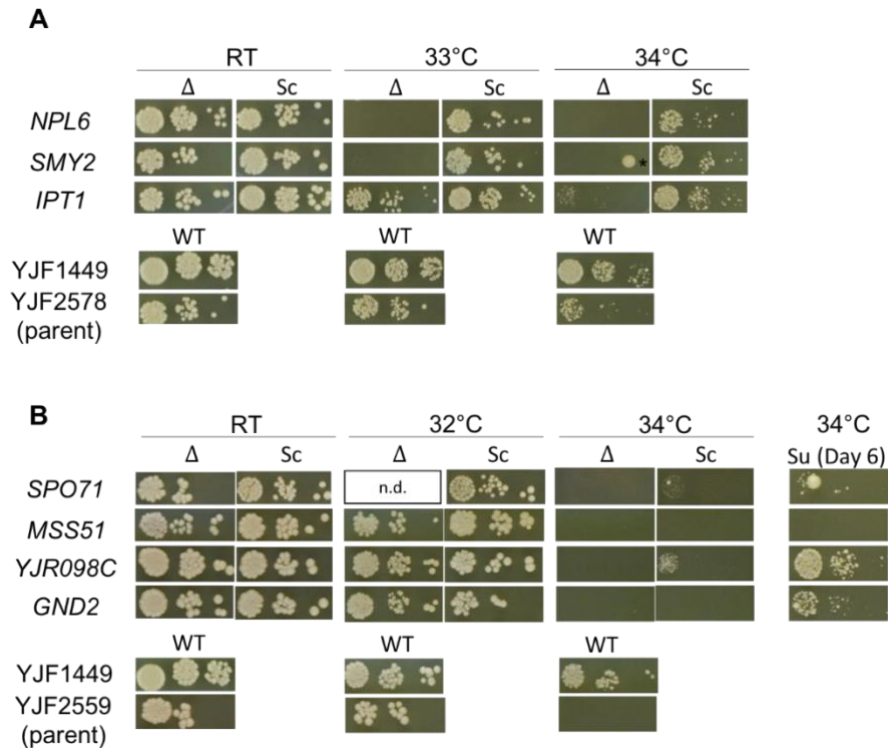


**Figure 5-10. Spot dilution of MoBY transformants and a non-transformant control.** Strains were spotted onto YPD+G418 plates with 1:4 serial dilution and incubated at room temperature or 33.5°C for 3 days. YJF2553 is a diploid *S. uvarum* strain with *KanMX* integrated into the *ho* locus. Other strains were YJF2602-derived MoBY transformants.

#### *Allele replacement of candidate genes*

Replacing the *S. uvarum* allele of candidate genes with the *S. cerevisiae* allele in *S. uvarum* would account for both plasmid burden and dosage effects. Via three routes of genetic manipulation (see Methods), I obtained allele replacements of 7 genes in total (Fig. 5-11). *NPL6* and *SMY2* were two of the top three candidates from the individual validation (scores of 6.0 and 5.4, respectively). *IPT1* was a candidate from the pooled competition but not confirmed by the plasmid assay. Consistent with their plasmid phenotype, knockouts of *NPL6*, *SMY2* but not *IPT1* showed heat sensitivity. However, none of the *S. cerevisiae* allele replacements improved *S. uvarum*'s growth at restrictive temperatures, compared to their parental strain YJF2578 (Fig. 5-11A). *SPO71*, *MSS51*, *YJR098C*, and *GND2* showed moderate effects in the plasmid assay (scores of 4.5, 3.3, 2.9, and 2.8, respectively). While allele replacements of *SPO71* and *YJR098C*

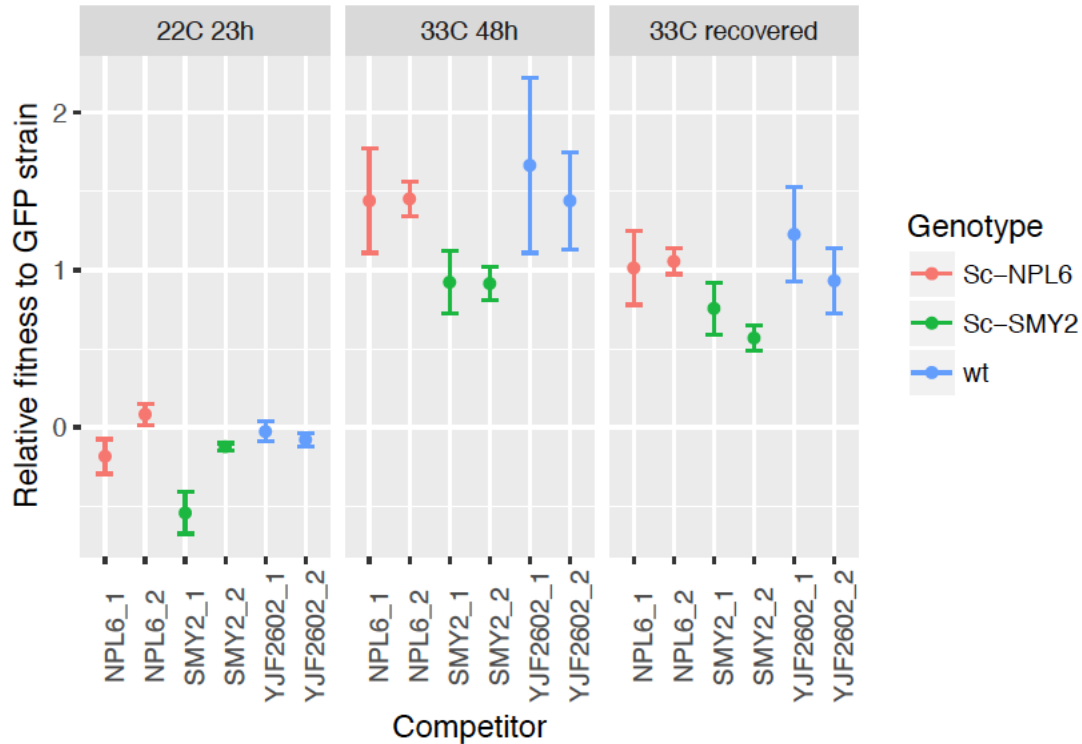
slightly enhanced the growth at 34°C compared to their parent strain YJF2559, the effect could be explained by a positional effect of the *ura3* locus, because integration of the *S. uvarum* alleles at the same locus showed similar effect (Fig. 5-11B).



**Figure 5-11. Allele replacements of 7 candidate genes. A.** YJF2578-derived allele replacements of *NPL6*, *SMY2* and *IPT1*. **B.** YJF2559-derived allele replacements of *SPO71*, *MSS51*, *YJR098C* and *GND2*. *NPL6* and *SMY2* were replaced at their endogenous loci; other genes were knocked out in *S. uvarum* (“Δ”) and the *S. cerevisiae* (“Sc”) allele was integrated at the *ura3* locus of the knockout strain. For *SPO71*, *MSS51*, *YJR098C* and *GND2*, the *S. uvarum* (“Su”) alleles were integrated at the same locus for control. Strains were plated with 1:10 serial dilution on YPD. Growth was after 3 days, with the exception of the *S. uvarum*-allele controls (6 days). Parental strains YJF2578 (*ura3-140*), YJF2559 (*ura3Δ::CORE-I-SceI*) and YJF1449 were

included for comparison. Note that both panels used the same picture of room temperature (“RT”) growth of YJF1449. n.d., not determined. WT, wild-type. Asterisk marks a contamination on the plate.

To measure possible small effects that cannot be seen on a plate, diploids that were homozygous for *S. cerevisiae* *NPL6* or *SMY2* were competed with a GFP-tagged strain at 22°C for 23h and 33°C for 48h (7-9 generations). The non-GFP competitors showed similar fitness to the GFP strain at 22°C and higher fitness at 33°C, but neither strain with allele replacements showed higher fitness than the wild-type *S. uvarum* control at high temperature (Fig. 5-12). The increase in relative fitness to GFP cells at 33°C could be explained by loss of GFP expression at this temperature, as the growth advantage of the competitors was reduced after a 24h-recovery at 22°C. Since there was no growth difference between the competitors and the GFP strain at 22°C (with the exception of *SMY2\_1*), the reduced growth advantage may be attributed to recovered GFP expression. Regardless, the growth competition was consistent with the plate assay in that *S. cerevisiae* *NPL6* or *SMY2* did not confer heat tolerance in *S. uvarum*.



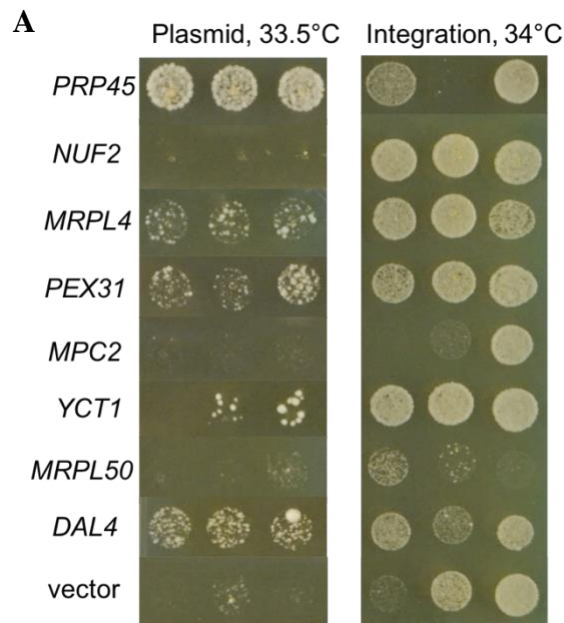
**Figure 5-12. Growth competition with a GFP strain.** Fitness was shown relative to the GFP strain (YJF2656), such that a fitness of 0 means no growth difference between the competitor and the GFP strain, a positive value means better growth of the competitor than the GFP strain and *vice versa*. Error bars represent standard error of 6 replicates. For each genotype two colonies were examined (“\_1” and \_2”).

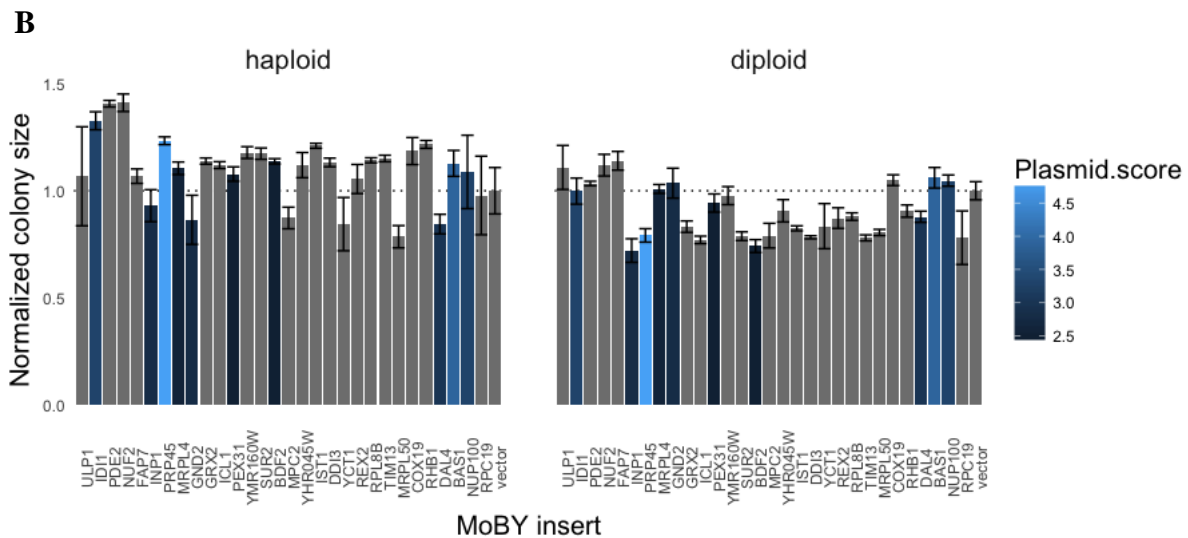
### *Integration of MoBY plasmids*

To screen the candidate genes in a high-throughput manner, I designed a strategy to integrate candidate MoBY plasmids into the *ura3* locus of *S. uvarum*, through stitching-PCR and gap repair. For an initial test I integrated 30 candidate plasmids that were positive hits from the pooled competition. At 32°C, all the transformants grew similarly (data not shown) and the growth at 34°C was variable across individual transformants of the same plasmid (Fig. 5-13A).

In particular, the variability of the vector control made it difficult to interpret the phenotype of candidate genes. Furthermore, there was a lack of correspondence between the plasmid and the integration phenotype, e.g. *PRP45* had a large effect on plasmid but not at the genomic locus (Fig. 5-13A).

The variation across transformants of the same gene could be potentially explained by secondary mutations or aneuploidy introduced during transformation. Secondary mutations are presumably recessive, and their effect should be eliminated in diploids. Therefore, I crossed the haploid transformants to wild-type *S. uvarum* and examined the growth of the haploids and the diploids arrayed on agar plates (Fig. 5-13B). All strains grew very slowly at 34-35°C; although several strains had higher mean growth than the vector control (e.g. *PDE2*, *NUF2*), the effects were either small or inconsistent between haploids and diploids. Again, there was no correspondence between the plasmid phenotype (bar color) and the integration phenotype (bar height, Fig. 5-13B).





**Figure 5-13. Integration of MoBY plasmids.** A. Plasmid and integration of MoBY transformants were grown at 4 days at 33.5°C and 3 days at 34°C, respectively. Eight candidate genes are shown as examples. Each spot was an independent transformant. B. Normalized colony size of integrated haploid transformants and their diploid derivatives, grown at 35°C for 9 days and 34°C for 5 days, respectively. Colonies sizes were normalized to the vector control. Error bars represent standard error of 5-12 colonies from 3 independent transformants. Growth scores at 33.5°C in the plasmid assay were shown for genes with a score greater than 2.5.

### *Plasmid burden*

The inconsistency between the plasmid and integration phenotypes point to plasmid burden as a potential confounding factor in our screen. However, the plasmid burden should not affect the results if it has a constant effect across genotypes and conditions. We thus examined the plasmid loss rate at 22°C and 30°C to see if the plasmid burden was affected by temperature (Table 5-2). The MoBY vector showed a high rate of plasmid loss at 22°C and the rate was



doubled at 30°C. The high rate of loss suggested a potential incompatibility between *S. cerevisiae* CEN sequence and *S. uvarum*. The plasmids with *S. cerevisiae* *UME1* and *MCO76* showed lower loss rates than the vector and their loss rates only increased by 1.7~1.8-fold at 30°C. These differences in plasmid retention among MoBY plasmids may explain the temperature effects of certain genes observed on G418 plates: the *MCO76* plasmid could confer heat tolerance presumably due to its low plasmid loss rate at high temperature.

Table 5-2. Plasmid loss per generation in MoBY transformants.

	22°C	30°C
Empty MoBY vector	14.5%	30.0%
<i>UME1</i> (weak temperature candidate)	4.7%	8.4%
<i>MCO76</i> (strong temperature candidate)	0.7%	1.2%

To support this idea, we further examined the distribution of autonomously replicating sequences (ARS) in the MoBY inserts. 7 of the 41 candidate genes validated by individual spot assays have ARS in their insert sequences, an enrichment compared to the background frequency (11 out of 80 sequenced plasmids). The ARS sequences might have helped resolve the replication problem of the MoBY plasmids and conferred better G418 growth. Taken together, plasmid burden is a significant confounding factor for growth phenotypes and should be taken into account in experimental design and data interpretation.

### *Heat shock of MoBY transformants*

The temperature-interactive plasmid burden may compromise screens based on growth competition. However, plasmid burden may have little or no effect if the selection does not depend on growth. In Part 1 of this chapter, I showed that heat tolerance correlates with survival at lethal heat shock (HS) temperatures in *S. cerevisiae* and *S. uvarum*. Therefore, we could select for MoBY transformants that *survive* better at heat shock conditions, instead of those showing a higher growth rate at restrictive temperatures. Exposure to 44°C for 1h typically kills 99% of *S. uvarum* cells. If an *S. cerevisiae* gene increases *S. uvarum*'s chance to survive the lethal heat shock, its frequency should increase in the transformant pool after a few HS-recovery cycles. Genes identified this way would have direct implications in species' divergence in HS survival and might also contribute to the divergence in high temperature growth, as the two traits may have some shared genetic basis.

As an initial test, the MoBY transformant pool of *S. uvarum* was subjected to 5 rounds of heat shock at 43-44°C, with 2-3 days' recovery growth at 22°C (4-10 generations) in between. Because spontaneous mutations can also contribute to the adaptation to HS, a single transformant of the MoBY vector was subjected to the same treatment, to provide a comparison between the effect of *S. cerevisiae* genes and that of beneficial mutations occurred in *S. uvarum*. Mock treatments at 22°C were used to control for selection during the recovery phase. After 5 rounds of selection, the transformant pool showed a great increase in HS survival (from 1.13% to 27.9%), while the survival rate of the vector control remained the same (~0.5%, Table 5-3), suggesting that there are likely *S. cerevisiae* genes in the pool that can increase HS survival. Mock treatments of both the pool and the vector control showed moderate increase in HS

survival (from ~1% to 5-8%), suggesting that the recovery phase might have also selected for some mutations that were beneficial to HS survival. There was also a difference between the mock-treated and the HS-treated cultures of the vector control, with the mock-treated showing better adaptation to HS (5.09% vs. 0.45%). This result could be explained by a difference in the population dynamics: although both groups were allowed to grow to saturation before the next round of HS, it took longer for the HS group to recover from the stress and resume dividing. The mock group was basically in continuous growth and might have selected more beneficial mutations. Also, the HS group had gone through more bottlenecks than the mock group, because the bottleneck effects were only controlled in the last two cycles in the mock group (see Methods).

Table 5-3. Heat shock survival rate after 5 HS-recovery cycles.

Sample	Treatment	2 <sup>nd</sup> cycle	5 <sup>th</sup> cycle
Transformant pool	Mock	1.13%	8.10%
Transformant pool	Heat shock		27.9%
Vector	Mock	0.47%	5.09%
Vector	Heat shock		0.45%

To test whether plasmid burden affect the HS results, I examined the HS survival of three strains with known plasmid burden. The survival rate of vector, *UME1* and *MCO76* was 3.63%, 1.89% and 2.23%, respectively, and did not correlate with their plasmid burden (Table 5-2). It suggests the HS strategy is less likely to be affected by plasmid burden than growth-based assays.

To test whether this strategy can be used to identify possible candidate genes, I isolated 5 colonies from the heat-shocked transformant pool and sequenced their barcodes (uptags). Two of the colonies did not have uptags; three colonies were mapped to MoBY plasmids of *PEX25*, *GOR1*, *SPP381*, respectively, none of which had known roles in thermotolerance (Table 5-4). Interestingly, four of the colonies showed high HS survival and one colony (#12) showed very low survival. One random colony from the transformant pool under mock treatment also showed a high survival rate (13.02%), while the average survival rate of the pool was only 8.10% (Table 5-3). These results suggest the adaptation to HS of the transformant pool may involve a diverse set of genes, rather than being dominated by one or a few MoBY plasmids.

Table 5-4. HS survival of individual colonies in the HS pool.

Source	ID	Survival	Gene (Uptag)	Gene function
HS pool	#7	13.67%	NA	NA
HS pool	#8	22.83%	NA	NA
HS pool	#9	15.93%	<i>PEX25</i>	Peripheral peroxisomal membrane peroxin
HS pool	#10	18.50%	<i>GOR1</i>	Glyoxylate reductase
HS pool	#12	<0.01%	<i>SPP381</i>	mRNA splicing factor
Mock pool	Random colony	13.02%	Not determined	NA

## Discussion

### *Technical challenges in cross-species genetics*

While the MoBY library allows a high-throughput approach to quantify fitness effects of thousands of *S. cerevisiae* genes simultaneously in an *S. uvarum* background, application of an *S. cerevisiae*-based tool to its related species requires extra caution regarding its compatibility.

While replication and maintenance of plasmid DNA is thought to be a minor burden in *S.*

*cerevisiae* (Karim et al. 2013), we found it to be a greater concern when *S. cerevisiae* CEN plasmids are introduced to *S. uvarum*. In our MoBY screen, although the pooled competition gave rise to a number of candidate genes and 41 of them were validated by spot assays of individual plasmid transformants, integration of the candidate genes did not recapitulate the plasmid phenotypes. Furthermore, the plasmid loss assay showed that the MoBY plasmids were lost frequently in *S. uvarum* and the loss rate interacted with temperature. The high loss rate was potentially due to incompatibility between *S. cerevisiae* CEN sequence and *S. uvarum* DNA replication machinery. The loss was also observed by other research groups, and the problem seemed to be caused specifically by CEN sequence rather than ARS (Maitreya Dunham, personal communication). Due to the difficulties in the sequence assembly of *S. uvarum* centromeres, it remained unclear what difference in the CEN sequence caused the incompatibility. An alternative experimental design would be to utilize the 2-micron MoBY library (Ho et al. 2009) for the screen since its replication does not depend on CEN. Also, the heat shock-based selection method described in this study could potentially eliminate the confounding effects of plasmid burden, although further validation of the HS-selected genes is needed before making a conclusion.

Another challenge in studying heat tolerance of *S. uvarum* is that the many standard genetic techniques such as CEN plasmids, GFP expression and transformation can be influenced by its heat sensitivity. The burden of CEN plasmid was worse at 30°C compared to 22°C and some inserts were able to mitigate this temperature effect, causing false positives in the screen. Competition with GFP-tagged strains is a common approach to determine a strain's fitness (Bergen et al. 2016). However, GFP expression was found to be unstable at 33°C in *S. uvarum*, increasing the noise of fitness measurements. Finally, the transformation of *S. uvarum* itself

could cause sensitivity at 34-35°C, although we used a modified heat shock temperature of 37°C in our protocol. For example, three transformants (YJF2559, YJF2560 and YJF2561) were picked when replacing *URA3* with CORE-I-SceI cassettes, but they showed different levels of sensitivity at 34°C, with YJF2559 being more sensitive than the original *S. uvarum* strain YJF1449. Such variability across transformants was also observed in other *S. uvarum* strains with genomic integrations. The variability could not be eliminated in backcrossed diploids, suggesting that aneuploidy rather than point mutations might be the underlying genetic factor. Although transformation was known to be mutagenic in *S. cerevisiae*, such high frequency of aneuploid-like transformants is unusual. Taken together, the *S. uvarum* cell might be a highly unstable system at temperatures above 30°C: plasmids get lost, gene expression is not robust, and any genetic disturbance might cause sensitive growth. From this prospective, the hybrids of *S. cerevisiae* and *S. uvarum* such as used in Chapter 3 would provide a more robust system for genetic manipulations. Alternatively, manipulations of *S. uvarum* could be performed in a careful way to control for such issues, e.g. sporulation after backcross to eliminate aneuploid transformants.

### *Genetic architecture of heat tolerance*

The plasmid burden could increase the false positive rate in the MoBY screen. However, the screen also assumes that there are single *S. cerevisiae* genes with detectable and additive effects on heat tolerance. From Chapter 3, the genetic architecture underlying thermal divergence between *S. cerevisiae* and *S. uvarum* is likely to be complex, potentially with many changes of small effects. A recent study on heat tolerance of *S. cerevisiae* and *S. paradoxus* supports this idea (Weiss et al. 2018): while 8 genes showed obvious allele differences in reciprocal

hemizygotes or *S. cerevisiae*, introducing their *S. cerevisiae* alleles into *S. paradoxus* only resulted in a marginal increase of heat tolerance. Therefore, it might be difficult to detect single genes of large effect when screening for additive effects if the underlying genetic architecture is highly complex.

## Chapter 6: Conclusion and future directions

The genetic architecture of phenotypic divergence between species is not well understood. Outstanding questions include the number of changes, the distribution of their effect sizes and the role of *cis*-regulatory changes and epistasis in phenotypic evolution. In the preceding chapters, I addressed these questions by studying the genetic basis of thermal divergence between *S. cerevisiae* and *S. uvarum*. My findings collectively support a polygenic architecture of evolution, with contributions from both *cis*-regulatory and coding changes. Particularly, multiple mitochondria-encoded genes contribute to evolution of heat and cold tolerance, pointing to the mitochondrial genome as a hotspot for yeast evolution. In this chapter, I will synthesize my findings across the chapters and propose a few future directions to further understand the genetic basis of thermal divergence.

### **A complex genetic architecture of interspecific differences**

A recurrent theme throughout my dissertation study is that evolution involves multiple small-to-moderate effect changes. In Chapter 3, I found a large effect of mitochondrial DNA (mtDNA) in heat- and cold-tolerant respiration, and it was caused by multiple genes, including a moderate effect of *COX1* and possibly small effects of genes like *COX2/COX3*. In Chapter 4, I dissected the effect of *CUP2* divergence on copper resistance and found that multiple changes contribute to the allele differences, with the proximal promoter region showing the largest effect. Furthermore, *HFA1* and *CUP2* showed moderate allele differences in heat and copper resistance, respectively, but neither the two genes, nor mtDNA, explained the total phenotypic difference between the two species. The surprisingly few genes uncovered by the genome-wide non-complementation screen also implies a lack of large effects among the genes screened. These



findings are consistent with a micromutationist view (in its weak version) that evolution occurs through accumulation of small-effect changes. However, we could not rule out large-effect genes contributing to thermal divergence, because our screen did not cover essential genes and genes whose effect are masked by other genes.

Through the dissection of *CUP2* allele differences, I found that both *cis*-regulatory and coding changes contributed to phenotypic divergence. The region with the largest effect in *CUP2* was the proximal promoter, consistent with an important role of *cis*-regulatory changes in evolution. *Cis*-regulatory divergence had prevalent effects on gene expression, but only affected temperature response in a small number of genes. Also, no pathway showed signatures of directional *cis*-regulatory evolution in a temperature-specific manner.

In the screens I mostly tested for additive effects, but epistasis was observed in a few cases. Mito-nuclear epistasis is likely to contribute to thermal divergence, as the *COX1* allele effect depended on nuclear background. The copper-resistant changes in *S. cerevisiae CUP2* were mostly additive, but there might be synergistic epistasis between the distal promoter and the C-terminal end of the coding sequence. In both cases, additional experiments would be needed to address the role of epistasis in phenotypic evolution.

My findings of mitochondrial DNA and *HFA1* only explained a fraction of the total difference in heat tolerance between *S. cerevisiae* and *S. uvarum*. How can we further understand the genetic basis of their thermal divergence? Mutagenesis and experimental evolution can be complementary approaches to the screens performed in this study. Screening for heat-sensitive mutants of *S. cerevisiae* × *S. uvarum* hybrids would allow identification of essential genes involved in thermal divergence, which were missing in our non-complementation screen.

Recently, Weiss et al. (2018) used transposon mutagenesis to screen for genes underlying thermal differences in *S. cerevisiae* and *S. paradoxus* and found eight housekeeping genes, highlighting the importance of screening essential genes. Experimental evolution of *S. cerevisiae* and *S. uvarum* hybrids or *S. uvarum* itself at high temperature could help identify heat tolerant mutations. Compared to mutagenesis, the experimental evolution method can detect small-effect mutations and interactions between successive mutations. Previous studies showed that evolving *S. cerevisiae* and *S. uvarum* hybrids at different temperatures had different genomic outcomes: the hybrids tend to lose *S. cerevisiae* alleles at low temperatures and *S. uvarum* alleles at high temperatures (Piotrowski et al. 2012; Smukowski Heil et al. 2018). Identification of loss-of-heterozygosity events in hybrids during experimental evolution could discover genes involved in thermal divergence.

The screens in Chapter 3 and 5 showed that thermal divergence is unlikely to be caused by single genes. QTL mapping with mitotic recombinants or F2 progeny of *S. cerevisiae* and *S. uvarum* may potentially identify multiple genes, although these approaches might suffer from insufficient statistical power and difficulties in fine-mapping as in traditional QTL studies. With the recent advance in genome editing technology, it might be possible to design high-throughput genetic manipulations, such as to introduce multiple *S. cerevisiae* genes into *S. uvarum*, or to knock out multiple *S. cerevisiae* alleles in the hybrids. However, it is unrealistic to exhaustively sample all the possible combinations between 5,000 genes with the combinatorial knockout/knockin method. Candidate pathways such as the heat shock signaling pathway, mitochondrial genes, or pathways under directional *cis*-regulatory evolution could be targeted instead.

Additionally, it would be helpful to decompose the growth phenotype into multiple cellular phenotypes and dissect their genetic basis. The idea is easy to understand if we consider thermal adaptation of plants and animals: for example, polar bears adapt to cold environment through multiple physiological changes such as thick layer of fat under the skin, small ears and short tails. Like yeast, the cold adaptation of polar bears is unlikely to be caused by a single gene; studying the fat layer would only reveal a subset of genes involved in the adaptation, but it could simplify the problem greatly. With yeast, the relevant cellular phenotypes may include respiration (Chapter 3), heat-shock signaling and membrane lipid composition (Chapter 5). Nutrient metabolism could also be important, as *S. uvarum* is sensitive to 30°C under glucose, phosphate and sulfate limitations but not *S. cerevisiae* (Smukowski Heil et al. 2017). For example, the non-complementation screen could be performed at phosphate-limited conditions at high temperature to screen for genes in the phosphate pathway that diverged in heat tolerance. For biochemical phenotypes such as membrane lipid composition and protein misfolding, collaboration with biochemists and cell biologists would allow us to describe and quantify the differences between the two species, and then use genetic methods to map the differences.

### **Mitochondrial genome as an evolutionary hotspot**

The yeast genetics field has recently seen increasing evidence of mitochondrial effects on phenotypic variation (Paliwal et al. 2014; Wolters et al. 2018), although pinpointing the causal changes in mitochondrial genome has been difficult. Across *Saccharomyces* species, prior research has shown that multiple mito-nuclear incompatibilities contribute to their reproductive isolation under the Dobzhansky-Muller model. My study showed that mitochondria-encoded genes also contribute to evolution of heat and cold tolerance in *Saccharomyces* species. Using an

innovative deletion mapping approach and biolistic transformation, I found multiple mitochondrial genes with small or moderate effects contribute to the divergence. The deletion mapping method can be potentially applied to other phenotypic effects of mtDNA within and between species. Additionally, I observed *ORF1*-mediated mtDNA recombination and identified an association between *ORF1* and mtDNA inheritance in the hybrids, which may have implications in both yeast mitochondrial genome evolution and the evolution of selfish elements.

The coincidence of “speciation genes” and “temperature genes” in the mitochondrial genome highlights its important role in yeast evolution. Do the Dobzhansky-Muller incompatibilities and thermal divergence share the same molecular basis? My results suggest that this is unlikely, but additional experiments are needed for a conclusion. The previously identified incompatible mitochondrial genes are *ATP9* and *COX1* (Lee et al. 2008; Chou et al. 2010). *ATP9* was not significantly associated with heat tolerance; *COX1* protein divergence showed a moderate effect in both heat and cold tolerance, but the previously reported incompatibility was between a *COX1* intron and a nuclear splicing factor *MRS1* (Chou et al. 2010). Future experiments could examine *COX1* splicing in reciprocal hemizygotes of *MRS1* in *S. cerevisiae* and *S. uvarum* hybrids, and how it is affected by mitotype and temperature. If the *S. cerevisiae* *MRS1-COX1* interaction is cold sensitive and the *S. uvarum* *MRS1-COX1* interaction is heat sensitive, it would suggest that thermal divergence might be a potential driver of the incompatibilities.

Perhaps the most surprising finding in the mitochondrial study is that the effect of *COX1* protein divergence depended on the nuclear background. *S. uvarum* Cox1p can function in the *S. cerevisiae* background at high temperature, despite a slight temperature-independent growth defect. However, *S. uvarum* Cox1p showed heat sensitivity in the hybrid background, suggesting

a dominant negative effect of the *S. uvarum* nuclear genome. Future experiments could determine what factors in the *S. uvarum* nuclear genome caused the temperature sensitivity. Proteins in the cytochrome *c* oxidase complex are strong candidates, since Cox1p is the core component of the complex and is in physical contact with many nuclear-encoded subunits (Mick et al. 2011). We could knock out the *S. uvarum* allele of candidate subunits in an *S. cerevisiae* × *S. uvarum* hybrid with *S. uvarum COX1* integrated into *S. cerevisiae* mitochondrial genome (generated in this study) and screen for rescue of heat sensitivity. Identification of the nuclear partner that co-evolved with Cox1p in *S. uvarum* genome might shed light on the genetic basis of mitochondria-mediated heat or cold tolerance.

The mitochondrial allele replacements found a negative effect of *COX1* and *COB* introns on high temperature growth in the hybrids. Also, intron-less *COB* mitigated the cold sensitivity in *S. cerevisiae*. Is intron processing a burden to yeast under heat or cold stress? What are the responsible nuclear factors? A suppressor screen could be performed to identify nuclear mutants that improve heat tolerance of *S. cerevisiae* × *S. uvarum* hybrids with intron-bearing *S. cerevisiae* mtDNA, either by random mutagenesis or systematically knocking out genes (presumably the *S. uvarum* alleles) involved in mitochondrial intron processing. Introns are known to co-evolve with nuclear factors (Rudan et al. 2018) and understanding their burden could provide insight on evolution of these selfish elements.

### **Trade-off between heat and cold tolerance**

My final thoughts on the genetic basis of thermal divergence would be a commentary and might not be based on a solid experimental evidence. However, I hope it could still provide some insights for future work. I argue that the trade-off between heat and cold tolerance needs to be

taken into account when studying the genetic basis of either phenotype. When focusing on the evolution of heat tolerance, I used to treat *S. uvarum* as a “heat sensitive” species and look for “heat sensitive” genes in its genome, which are more or less assumed as loss-of-function changes. However, as the non-complementation screen and gene expression study showed, *S. uvarum* genes/promoters are usually not heat sensitive. Instead, dominant negative effects of *S. uvarum* genes were recurrently observed in the hybrid: the hybrids generally grow slower than *S. cerevisiae* at 37-39°C; the gene expression profile reflects a dominant trans-acting effect of *S. uvarum* genome at 22°C; the *COX1* effect was only present in the hybrid background but not *S. cerevisiae*. It seems that *S. uvarum* gene products cause problems, rather than stop working, at high temperature.

To better understand the genetic basis of thermal divergence, we may have to consider the cryotolerance of *S. uvarum*. In thermal adaptation, proteins or gene-gene interactions may be under biochemical constraints so that they can only function at high or low temperatures, e.g. enzymes have optimum temperatures. The heat sensitivity of *S. uvarum* may result from a trade-off between heat and cold tolerance, and some cold-adapted genes might cause problems at high temperature (e.g. misfolding, mis-binding, etc.). In Chapter 3, I found that the *S. cerevisiae* mitotype is heat tolerant and the *S. uvarum* mitotype is cold tolerant, but neither can tolerate both heat and cold temperatures, potentially suggesting a trade-off. This also happens at the single gene level where *S. cerevisiae* Cox1p conferred heat tolerance and *S. uvarum* Cox1p conferred cold tolerance. The lipid composition discussed in Chapter 5 is another example: the level of unsaturated fatty acids in cell membranes is high in cryotolerant yeasts and low in thermotolerant yeasts (Arthur and Watson 1976), suggesting the heat- and cold-adaptations are mutually exclusive. Consistent with the trade-off view, a recent study found that *PHO84*, a high affinity

phosphate transporter, showed reciprocal loss-of-heterozygosity (LOH) in experimentally evolved *S. cerevisiae* and *S. uvarum* hybrids, with the *S. cerevisiae* allele being lost at low temperature and the *S. uvarum* allele being lost at high temperature (Smukowski Heil et al. 2018). Although the authors suggest mutations with such antagonistic pleiotropic effects were rare in their evolved lines, a systematic examination may be needed to evaluate the prevalence of heat-cold trade-offs between the two species. The frequent LOH events in the hybrids are also consistent with a dominant effect of the *S. uvarum* genome at high temperature. As such, a most radical, if not crazy, future experiment would be to transform a library of *S. uvarum* genes into *S. cerevisiae* and screen for dominant negative effects, assuming the alleles causing burden at high temperatures are cold-adapted (and heat sensitive).

Although all organisms are adapted to their thermal environment and many heat- or cold-related traits are described for animals and plants (such as the fat layer of polar bears), little is known about the genetic constraints in the trade-offs between two related species with different thermal profiles. Therefore, the prevalence and importance of genetic trade-offs between heat- and cold-adaptation are still open questions. However, these concerns could lead to a reflection that we should consider a trait's evolutionary history when researching its genetic basis. For example, copper resistance is considered as a derived character in the *S. cerevisiae* lineage, and *S. uvarum* may represent the ancestral state. In this case, we tend to expect that introducing the copper-resistant changes of *S. cerevisiae* into *S. uvarum* can increase its resistance, and the effects of these changes do not depend on a hybrid background. The ancestral state of temperature tolerance in the *Saccharomyces* lineage is not clear; it is possible that both *S. cerevisiae*'s heat tolerance and *S. uvarum*'s cold tolerance are derived, and evolution of one trait constrains the other. In this case, the genetic architecture is more complicated, and we might not

assume *S. cerevisiae* genes have additive effects in *S. uvarum* on heat tolerance. Hopefully, the experiments proposed earlier in this chapter would reveal more genes underlying thermal divergence and help understand the relationship between heat and cold tolerance in yeast.



## References

Albertin W, da Silva T, Rigoulet M, Salin B, Masneuf-Pomarede I, de Vienne D, Sicard D, Bely M, Marullo P. 2013. The mitochondrial genome impacts respiration but not fermentation in interspecific *Saccharomyces* hybrids. *PLoS One*. 8(9):e75121.

doi:10.1371/journal.pone.0075121.

Anders S, Huber W, Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M, Mortazavi A, et al. 2010. Differential expression analysis for sequence count data. *Genome Biol*. 11(10):R106. doi:10.1186/gb-2010-11-10-r106.

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 31(2):166–169. doi:10.1093/bioinformatics/btu638.

Arthur H, Watson K. 1976. Thermal adaptation in yeast: growth temperatures, membrane lipid, and cytochrome composition of psychrophilic, mesophilic, and thermophilic yeasts. *J Bacteriol*. 128(1):56–68.

Auesukaree C, Damnernsawad A, Kruatrachue M, Pokethitiyook P, Boonchird C, Kaneko Y, Harashima S. 2009. Genome-wide identification of genes involved in tolerance to various environmental stresses in *Saccharomyces cerevisiae*. *J Appl Genet*. 50(3):301–10.

doi:10.1007/BF03195688.

Baker EP, Peris D, Moriarty R V, Li XC, Fay JC, Hittinger CT. 2018. Mitochondrial DNA and temperature tolerance in lager yeasts. *bioRxiv*.:1–18. doi:10.1088/0957-4484/21/3/035701.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko

SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* 19(5):455–477.  
doi:10.1089/cmb.2012.0021.

Barnett JA. 2007. A history of research on yeasts 10: foundations of yeast genetics. *Yeast.* 24(10):799–845. doi:10.1002/yea.1513.

Beardsley PM, Schoenig SE, Whittall JB, Olmstead RG. 2004. Patterns of evolution in western North American *Mimulus* (Phrymaceae). *Am J Bot.* 91(3):474–489. doi:10.3732/ajb.91.3.474.

Becker B, Feller A, El Alami M, Dubois E, Piérard A. 1998. A nonameric core sequence is required upstream of the LYS genes of *Saccharomyces cerevisiae* for Lys14p-mediated activation and apparent repression by lysine. *Mol Microbiol.* 29(1):151–163. doi:10.1046/j.1365-2958.1998.00916.x.

Beney L, Gervais P. 2001. Influence of the fluidity of the membrane on the response of microorganisms to environmental stresses. *Appl Microbiol Biotechnol.* 57(1–2):34–42. doi:10.1007/s002530100754.

Bergen AC, Olsen GM, Fay JC. 2016. Divergent MLS1 Promoters Lie on a Fitness Plateau for Gene Expression. *Mol Biol Evol.* 33(5):1270–1279. doi:10.1093/molbev/msw010.

De Boer CG, Hughes TR. 2012. YeTFaSCo: A database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res.* 40(D1):169–179. doi:10.1093/nar/gkr993.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Bonnefoy N, Fox TD. 2000. In vivo analysis of mutated initiation codons in the mitochondrial COX2 gene of *Saccharomyces cerevisiae* fused to the reporter gene ARG8 reveals lack of downstream reinitiation. *Mol Gen Genet.* 262(6):1036–1046. doi:10.1007/PL00008646.

Bonnefoy N, Fox TD. 2001. Genetic transformation of *Saccharomyces cerevisiae* mitochondria. In: *Methods in Cell Biology.* Vol. 65. p. 381–396.

Bordonné R, Dirheimer G, Martin RP. 1988. Expression of the *oxi1* and maturase-related RF1 genes in yeast mitochondria. *Curr Genet.* 13(3):227–233. doi:10.1007/BF00387768.

Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, Holmes I, Pachter L. 2009. Fast Statistical Alignment. Siepel A, editor. *PLoS Comput Biol.* 5(5):e1000392. doi:10.1371/journal.pcbi.1000392.

Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. 2014. Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet.* 10(6). doi:10.1371/journal.pgen.1004410.

Buchman C, Skroch P, Welch J, Fogel S, Karin M. 1989. The CUP2 gene product, regulator of yeast metallothionein expression, is a copper-activated DNA-binding protein. *Mol Cell Biol.* 9(9):4091–4095. doi:10.1128/MCB.9.9.4091.

Bullard JH, Mostovoy Y, Dudoit S, Brem RB. 2010. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc Natl Acad Sci U S A.* 107(11):5058–63. doi:10.1073/pnas.0912959107.

Burt A, Koufopanou V. 2004. Homing endonuclease genes: the rise and fall and rise again of a

selfish element. *Curr Opin Genet Dev.* 14(6):609–615. doi:10.1016/j.gde.2004.09.010.

Burton RS, Barreto FS. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol.* 21(20):4942–57. doi:10.1111/mec.12006.

Carratù L, Franceschelli S, Pardini CL, Kobayashi GS, Horvath I, Vigh L, Maresca B. 1996. Membrane lipid perturbation modifies the set point of the temperature of heat shock response in yeast. *Proc Natl Acad Sci U S A.* 93(9):3870–3875. doi:10.1073/pnas.93.9.3870.

Carroll SB. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell.* 101(6):577–80.

Caudy AA, Guan Y, Jia Y, Hansen C, DeSevo C, Hayes AP, Agee J, Alvarez-Dominguez JR, Arellano H, Barrett D, et al. 2013. A New System for Comparative Functional Genomics of *Saccharomyces* Yeasts. *Genetics.* 195(1):275–287. doi:10.1534/genetics.113.152918.

Chang J, Zhou Y, Hu X, Lam L, Henry C, Green EM, Kita R, Kobor MS, Fraser HB. 2013. The Molecular Mechanism of a Cis-Regulatory Adaptation in Yeast. *PLoS Genet.* 9(9):1–8. doi:10.1371/journal.pgen.1003813.

Chang SL, Lai HY, Tung SY, Leu JY. 2013. Dynamic Large-Scale Chromosomal Rearrangements Fuel Rapid Adaptation in Yeast Populations. *PLoS Genet.* 9(1). doi:10.1371/journal.pgen.1003232.

Chen K, van Nimwegen E, Rajewsky N, Siegal ML. 2010. Correlating gene expression variation with cis-regulatory polymorphism in *Saccharomyces cerevisiae*. *Genome Biol Evol.* 2(1):697–707. doi:10.1093/gbe/evq054.

Chou J-Y, Hung Y-S, Lin K-H, Lee H-Y, Leu J-Y. 2010. Multiple molecular mechanisms cause reproductive isolation between three yeast species. Noor MAF, editor. PLoS Biol.

8(7):e1000432. doi:10.1371/journal.pbio.1000432.

Craig EA, Jacobsen K. 1984. Mutations of the heat inducible 70 kilodalton genes of yeast confer temperature sensitive growth. Cell. 38(3):841–849. doi:10.1016/0092-8674(84)90279-4.

Darwin CR. 1859. On the Origin of Species. London: Murray.

Dashko S, Liu P, Volk H, Butinar L, Piškur J, Fay JC. 2016. Changes in the relative abundance of two *Saccharomyces* species from oak forests to wine fermentations. Front Microbiol.

7(FEB):1–12. doi:10.3389/fmicb.2016.00215.

Davidson JF, Schiestl RH. 2001. Mitochondrial respiratory electron carriers are involved in oxidative stress during heat stress in *Saccharomyces cerevisiae*. Mol Cell Biol. 21(24):8483–9.

doi:10.1128/MCB.21.24.8483-8489.2001.

Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG. 2003. Engineering evolution to study speciation in yeasts. Nature. 422(6927):68–72. doi:10.1038/nature01418.

Dimitrov LN, Brem RB, Kruglyak L, Gottschling DE. 2009. Polymorphisms in Multiple Genes Contribute to the Spontaneous Mitochondrial Genome Instability of *Saccharomyces cerevisiae*

S288C Strains. Genetics. 183(1):365–383. doi:10.1534/genetics.109.104497.

Ding MG, Butler CA, Saracco SA, Fox TD, Godard F, di Rago J, Trumpower BL. 2009. Chapter 27 An Improved Method for Introducing Point Mutations into the Mitochondrial Cytochrome b Gene to Facilitate Studying the Role of Cytochrome b in the Formation of Reactive Oxygen

Species. In: *Methods in Enzymology*. Vol. 456. p. 491–506.

Dohmen RJ, Strasser AWM, Höner CB, Hollenberg CP. 1991. An efficient transformation procedure enabling long-term storage of competent cells of various yeast genera. *Yeast*. 7(7):691–692.

Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, Fay JC. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet*. 4(8).  
doi:10.1371/journal.pgen.1000183.

Duveau F, Toubiana W, Wittkopp PJ. 2017. Fitness Effects of Cis-Regulatory Variants in the *Saccharomyces cerevisiae* TDH3 Promoter. *Mol Biol Evol*. 34(11):2908–2912.  
doi:10.1093/molbev/msx224.

Eleutherio ECA, Araujo PS, Panek AD. 1993. Protective Role of Trehalose during Heat Stress in *Saccharomyces cerevisiae*. *Cryobiology*. 30(6):591–596. doi:10.1006/cryo.1993.1061.

Engle EK, Fay JC. 2012. Divergence of the Yeast Transcription Factor FZF1 Affects Sulfite Resistance. Malik HS, editor. *PLoS Genet*. 8(6):e1002763. doi:10.1371/journal.pgen.1002763.

Engle EK, Fay JC. 2013. ZRT1 Harbors an Excess of Nonsynonymous Polymorphism and Shows Evidence of Balancing Selection in *Saccharomyces cerevisiae*. *G3 (Bethesda)*. 3(April):665–673. doi:10.1534/g3.112.005082.

Fay JC, McCullough HL, Sniegowski PD, Eisen MB, Schadt E, Monks S, Drake T, Lusk A, Chen N, Colinayo V, et al. 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol*. 5(4):R26. doi:10.1186/gb-

2004-5-4-r26.

Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity (Edinb)*. 100(2):191–9. doi:10.1038/sj.hdy.6801000.

Fear JM, Leon-Novelo LG, Morse AM, Gerken AR, Van Lehmann K, Tower J, Nuzhdin S V., McIntyre LM. 2016. Buffering of genetic regulatory networks in *Drosophila melanogaster*. *Genetics*. 203(3):1177–1190. doi:10.1534/genetics.116.188797.

Fischer G, Neuvéglise C, Durrens P, Gaillardin C, Dujon B. 2001. Evolution of gene order in the genomes of two related yeast species. *Genome Res*. 11(12):2009–19. doi:10.1101/gr.212701.

Fisher RA. 1930. *The genetical theory of natural selection*.

Fogel S, Welch JW. 1982. Tandem gene amplification mediates copper resistance in yeast. *Proc Natl Acad Sci USA*. 79(September):5342–5346. doi:10.1073/pnas.79.17.5342.

Frankel N, Erezyilmaz DF, Mcgregor AP, Wang S, Stern DL. 2011. Morphological evolution caused by many subtle-effect substitutions in a transcriptional enhancer. *Nature*. 474(7353):598–603. doi:10.1038/nature10200.Morphological.

Fraser HB. 2011. Genome-wide approaches to the study of adaptive gene expression evolution: Systematic studies of evolutionary adaptations involving gene expression will allow many fundamental questions in evolutionary biology to be addressed. *BioEssays*. 33(6):469–477. doi:10.1002/bies.201000094.

Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, Mehrabian M, Schadt EE. 2011. Systematic Detection of Polygenic cis-Regulatory Evolution. Nachman MW, editor. *PLoS Genet*.

7(3):e1002023. doi:10.1371/journal.pgen.1002023.

Fraser HB, Levy S, Chavan A, Shah HB, Perez JC, Zhou Y, Siegal ML, Sinha H. 2012. Polygenic cis-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Res.* 22(10):1930–1939. doi:10.1101/gr.134080.111.

Fraser HB, Moses AM, Schadt EE. 2010. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc Natl Acad Sci U S A.* 107(7):2977–82. doi:10.1073/pnas.0912245107.

Fritsch ES, Chabbert CD, Klaus B, Steinmetz LM. 2014. A Genome-Wide Map of Mitochondrial DNA Recombination in Yeast. *Genetics.* 198(2):755–771. doi:10.1534/genetics.114.166637.

Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB. 2004. Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol.* 2(12):e398. doi:10.1371/journal.pbio.0020398.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Mol Biol Cell.* 11(12):4241–4257. doi:10.1091/mbc.11.12.4241.

Gerke J, Lorenz K, Cohen B. 2009. Genetic interactions between transcription factors cause natural variation in yeast. *Science.* 323(5913):498–501. doi:10.1126/science.1166426.

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison III CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* 6:343.

Gibson G. 2008. The environmental contribution to gene expression profiles. *Nat Rev Genet.*



9(8):575–81. doi:10.1038/nrg2383.

Gietz RD, Schiestl RH, Willems AR, Woods RA. 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*. 11(4):355–360.

doi:10.1002/yea.320110408.

Gonçalves P, Valério E, Correia C, de Almeida JMGCF, Sampaio JP. 2011. Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species.

*PLoS One*. 6(6):e20739. doi:10.1371/journal.pone.0020739.

Graden JA, Posewitz MC, Simon JR, George GN, Pickering IJ, Winge DR. 1996. Presence of a copper(I)-thiolate regulatory domain in the copper-activated transcription factor Amt1.

*Biochemistry*. 35(46):14583–9. doi:10.1021/bi961642v.

Grishkevich V, Yanai I. 2013. The genomic determinants of genotype × environment interactions in gene expression. *Trends Genet*. 29(8):479–487. doi:10.1016/j.tig.2013.05.006.

Gruber JD, Vogel K, Kalay G, Wittkopp PJ. 2012. Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *saccharomyces cerevisiae*: Frequency, effects, and dominance. *PLoS Genet*. 8(2). doi:10.1371/journal.pgen.1002497.

Grundberg E, Adoue V, Kwan T, Ge B, Duan QL, Lam KCL, Koka V, Kindmark A, Weiss ST, Tantisira K, et al. 2011. Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet*. 7(1). doi:10.1371/journal.pgen.1001279.

Hagman A, Säll T, Compagno C, Piskur J. 2013. Yeast “Make-Accumulate-Consume” Life Strategy Evolved as a Multi-Step Process That Predates the Whole Genome Duplication.

Fairhead C, editor. PLoS One. 8(7):e68734. doi:10.1371/journal.pone.0068734.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Res.* 41(13):e129–e129. doi:10.1093/nar/gkt371.

Hartl DL, Clark AG, Clark AG. 1997. Principles of population genetics. Sinauer associates Sunderland.

He F, Arce A, Schmitz G, Beyer A, Koornneef M, Novikova P, de Meaux J. 2016. The footprint of polygenic adaptation on stress-responsive cis-regulatory divergence in the Arabidopsis genus. *Mol Biol Evol.* 33(8):1–34. doi:10.1093/molbev/msw096.

Hertz GZ, Stormo GD. 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics.* 15(7–8):563–577. doi:10.1093/bioinformatics/15.7.563.

Ho CH, Magtanong L, Barker SL, Gresham D, Nishimura S, Natarajan P, Koh JLY, Porter J, Gray CA, Andersen RJ, et al. 2009. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol.* 27(4):369–77. doi:10.1038/nbt.1534.

Hodgins-Davis A, Townsend JP. 2009. Evolving gene expression: from G to E to G×E. *Trends Ecol Evol.* 24(12):649–658. doi:10.1016/j.tree.2009.06.011.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution.* 61(5):995–1016. doi:10.1111/j.1558-5646.2007.00105.x.

Hoja U, Marthol S, Hofmann J, Stegner S, Schulz R, Meier S, Greiner E, Schweizer E. 2004.

HFA1 encoding an organelle-specific acetyl-CoA carboxylase controls mitochondrial fatty acid synthesis in *Saccharomyces cerevisiae*. *J Biol Chem.* 279(21):21779–21786.

doi:10.1074/jbc.M401071200.

Hunter N, Chambers SR, Louis EJ, Borts RH. 1996. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J.* 15(7):1726–1733.

doi:10.1038/nature05099.

Jacquier A, Dujon B. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell.* 41(2):383–394. doi:10.1016/S0092-

8674(85)80011-8.

Jarolim S, Ayer A, Pillay B, Gee AC, Phrakaysone A, Perrone GG, Breitenbach M, Dawes IW.

2013. *Saccharomyces cerevisiae* genes involved in survival of heat shock. *G3 (Bethesda).*

3(12):2321–33. doi:10.1534/g3.113.007971.

Jhuang H, Lee H, Leu J. 2017. Mitochondrial–nuclear co-evolution leads to hybrid incompatibility through pentatricopeptide repeat proteins. *EMBO Rep.* 18(1):87–101.

doi:10.15252/embr.201643311.

Jin YH, Dunlap PE, McBride SJ, Al-Refai H, Bushel PR, Freedman JH. 2008. Global transcriptome and deletome profiles of yeast exposed to transition metals. *PLoS Genet.* 4(4).

doi:10.1371/journal.pgen.1000053.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ.

2016. HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-Throughput Sequencing Reads Using Target Enrichment. *Appl Plant Sci.* 4(7):1600016.

doi:10.3732/apps.1600016.

Jones CD. 1998. The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. *Genetics*. 149(4):1899–908.

Kahm M, Hasenbrink G, Lichtenberg-frate H, Ludwig J, Kschischo M. 2010. Grofit: Fitting biological growth curves. *J Stat Softw*. 33(7):1–21. doi:10.1038/npre.2010.4508.1.

Karim AS, Curran KA, Alper HS. 2013. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res*. 13(1):107–116. doi:10.1111/1567-1364.12016.

Kastaniotis AJ, Autio KJ, Kerätär JM, Monteuuis G, Mäkelä AM, Nair RR, Pietikäinen LP, Shvetsova A, Chen Z, Hiltunen JK. 2017. Mitochondrial fatty acid synthesis, fatty acids and mitochondrial physiology. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 1862(1):39–48. doi:10.1016/j.bbalip.2016.08.011.

Kawahara Y, Imanishi T. 2007. A genome-wide survey of changes in protein evolutionary rates across four closely related species of *Saccharomyces sensu stricto* group. *BMC Evol Biol*. 7:1–13. doi:10.1186/1471-2148-7-9.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28(12):1647–1649. doi:10.1093/bioinformatics/bts199.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of

yeast species to identify genes and regulatory elements. *Nature*. 423(6937):241–54.  
doi:10.1038/nature01644.

Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S, Weinberger A, Alon U, Milo R, Segal E. 2016. Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell*. 166(5):1282–1294.e18. doi:10.1016/j.cell.2016.07.024.

Kim HS, Huh J, Riles L, Reyes A, Fay JC. 2012. A Noncomplementation Screen for Quantitative Trait Alleles in *Saccharomyces cerevisiae*. *G3; Genes|Genomes|Genetics*. 2(7):753–760. doi:10.1534/g3.112.002550.

Kvitek DJ, Will JL, Gasch AP. 2008. Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. *PLoS Genet*. 4(10):e1000223. doi:10.1371/journal.pgen.1000223.

Landry CR, Oh J, Hartl DL, Cavalieri D. 2006. Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene*. 366(2):343–351. doi:10.1016/j.gene.2005.10.042.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359. doi:10.1038/nmeth.1923.

Leducq JB, Henault M, Charron G, Nielly-Thibault L, Terrat Y, Fiumera HL, Shapiro BJ, Landry CR. 2017. Mitochondrial recombination and introgression during speciation by hybridization. *Mol Biol Evol*. 34(8):1947–1959. doi:10.1093/molbev/msx139.

Lee H-Y, Chou J-Y, Cheong L, Chang N-H, Yang S-Y, Leu J-Y. 2008. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*.

135(6):1065–73. doi:10.1016/j.cell.2008.10.047.

Levo M, Zalckvar E, Sharon E, Dantas Machado AC, Kalma Y, Lotam-Pompan M, Weinberger A, Yakhini Z, Rohs R, Segal E. 2015. Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* 25(7):1018–29. doi:10.1101/gr.185033.114.

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature.* 519(7542):181–186. doi:10.1038/nature14279.

Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. *Science* (80- ). 352(6287):837–840. doi:10.1126/science.aae0568.

Li Y, Álvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, Hazendonk E, Prins P, Plasterk RHA, Jansen RC, et al. 2006. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* 2(12):2155–2161. doi:10.1371/journal.pgen.0020222.

Liti G, Barton DBH, Louis EJ. 2006. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics.* 174(2):839–50. doi:10.1534/genetics.106.062166.

Liu Z, Butow R a. 2006. Mitochondrial retrograde signaling. *Annu Rev Genet.* 40:159–185. doi:10.1146/annurev.genet.40.110405.090613.

Lopes CA, Barrio E, Querol A. 2010. Natural hybrids of *S. cerevisiae* × *S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. *FEMS Yeast Res.* 10(4):412–421. doi:10.1111/j.1567-1364.2010.00614.x.

- López-maury L, Marguerat S, Bähler J. 2008. Tuning gene expression to changing to evolutionary adaptation. *Nat Rev Genet.* 9(July):583–593. doi:10.1038/nrg2398.
- Mahmud SA, Hirasawa T, Shimizu H. 2010. Differential importance of trehalose accumulation in *Saccharomyces cerevisiae* in response to various environmental stresses. *J Biosci Bioeng.* 109(3):262–266. doi:10.1016/j.jbiosc.2009.08.500.
- Martin A, Orgogozo V. 2013. The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution (N Y).* 67(5):1235–1250. doi:10.1111/evo.12081.
- Martin HC, Roop JJ, Schraiber JG, Hsu TY, Brem RB. 2012. Evolution of a membrane protein regulon in *Saccharomyces*. *Mol Biol Evol.* 29(7):1747–56. doi:10.1093/molbev/mss017.
- Massey JH, Wittkopp PJ. 2016. *The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species.* 1st ed. Elsevier Inc.
- McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature.* 448(7153):587–590. doi:10.1038/nature05988.
- McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* 20(6):816–825. doi:10.1101/gr.102491.109.
- Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. 2015. Selection on noise constrains variation in a eukaryotic promoter. *Nature.* 521(7552):344–347. doi:10.1038/nature14244.

Mick DU, Fox TD, Rehling P. 2011. Inventory control: Cytochrome c oxidase assembly regulates mitochondrial translation. *Nat Rev Mol Cell Biol.* 12(1):14–20. doi:10.1038/nrm3029.

Mortimer RK. 2000. Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Res.* 10(4):403–409. doi:10.1101/gr.10.4.403.

Nagy O, Nuez I, Savisaar R, Peluffo AE, Yassin A, Lang M, Stern DL, Matute DR, David JR, Courtier-Orgogozo V. 2018. Correlated Evolution of two Copulatory Organs via a Single Cis-Regulatory Nucleotide Change. *Bioarxiv.*

Naranjo S, Smith JD, Artieri CG, Zhang M, Zhou Y, Palmer ME, Fraser HB. 2015. Dissecting the Genetic Basis of a Complex cis-Regulatory Adaptation. *PLoS Genet.* 11(12):1–19. doi:10.1371/journal.pgen.1005751.

Okuno M, Kajitani R, Ryusui R, Morimoto H, Kodama Y, Itoh T. 2016. Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res.* 23(1):67–80. doi:10.1093/dnares/dsv037.

Orr H. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution (N Y).* 52(4):935–949.

Orr H, Coyne J. 1992. The genetics of adaptation: a reassessment. *Am Nat.* 140(5):725–742.

Orr HA. 2001. The genetics of species differences. *Trends Ecol Evol.* 16(7):343–350.

Paget CM, Schwartz JM, Delneri D. 2014. Environmental systems biology of cold-tolerant phenotype in *Saccharomyces* species adapted to grow at different temperatures. *Mol Ecol.* 23(21):5241–5257. doi:10.1111/mec.12930.



Paliwal S, Fiumera AC, Fiumera HL. 2014. Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics*. 198(3):1251–65. doi:10.1534/genetics.114.168575.

Peichel CL, Marques DA. 2016. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos Trans R Soc London B Biol Sci*. 372(1713).

Perez-Martinez X, Broadley SA, Fox TD. 2003. Mss51p promotes mitochondrial Cox1p synthesis and interacts with newly synthesized Cox1p. *EMBO J*. 22(21):5951–61. doi:10.1093/emboj/cdg566.

Peris D, Arias A, Orlić S, Belloch C, Pérez-Través L, Querol A, Barrio E. 2017. Mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species. *Mol Phylogenet Evol*. 108:49–60. doi:10.1016/j.ympev.2017.02.008.

Peris D, Moriarty R V, Alexander WG, Baker E, Sylvester K, Sardi M, Langdon QK, Libkind D, Wang Q-M, Bai F-Y, et al. 2017. Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. *Biotechnol Biofuels*. 10(1):78. doi:10.1186/s13068-017-0763-7.

Petitjean M, Teste M-A, François JM, Parrou J-L. 2015. Yeast Tolerance to Various Stresses Relies on the Trehalose-6P Synthase (Tps1) Protein, Not on Trehalose. *J Biol Chem*. 290(26):16177–16190. doi:10.1074/jbc.M115.653899.

Piatkowska EM, Naseeb S, Knight D, Delneri D. 2013. Chimeric Protein Complexes in Hybrid Species Generate Novel Phenotypes. *PLoS Genet*. 9(10):e1003836. doi:10.1371/journal.pgen.1003836.

Picazo C, Gamero-Sandemetrio E, Orozco H, Albertin W, Marullo P, Matallana E, Aranda A. 2014. Mitochondria inheritance is a key factor for tolerance to dehydration in wine yeast production. *Lett Appl Microbiol.* 60(3):217–222. doi:10.1111/lam.12369.

Piotrowski JS, Nagarajan S, Kroll E, Stanbery A, Chiotti KE, Kruckeberg AL, Dunn B, Sherlock G, Rosenzweig F. 2012. Different selective pressures lead to different genomic outcomes as newly-formed hybrid yeasts evolve. *BMC Evol Biol.* 12(1):46. doi:10.1186/1471-2148-12-46.

Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature.* 440(7087):1050–1053. doi:10.1038/nature04597.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. *Nature.* 457:843.

Rak M, Tetaud E, Godard F, Sagot I, Salin B, Duvezin-Caubet S, Slonimski PP, Rytka J, Di Rago JP. 2007. Yeast cells lacking the mitochondrial gene encoding the ATP synthase subunit 6 exhibit a selective loss of complex IV and unusual mitochondrial morphology. *J Biol Chem.* 282(15):10853–10864. doi:10.1074/jbc.M608692200.

Rak M, Tzagoloff A. 2009. F1-dependent translation of mitochondrially encoded Atp6p and Atp8p subunits of yeast ATP synthase. *Proc Natl Acad Sci U S A.* 106(44):18509–14. doi:10.1073/pnas.0910351106.

Rasband W. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <https://imagej.nih.gov/ij/>, 1997-2016.

Rockman M V. 2011. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* (N Y). 66(1):1–17. doi:10.1111/j.1558-5646.2011.01486.x.

Romero I, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet.* 13(7):505–516. doi:citeulike-article-id:10800937 doi:10.1038/nrg3229.

Roop JI, Chang KC, Brem RB. 2016. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature.*:1–14. doi:10.1038/nature16938.

Rudan M, Dib PB, Musa M, Kanunnikau M, Sobočanec S, Rueda D, Warnecke T, Kriško A. 2018. Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. *Elife.* 7:1–17. doi:10.7554/eLife.35330.

Sadhu MJ, Bloom JS, Day L, Kruglyak L. 2016. CRISPR-directed mitotic recombination enables Mapping Without Crosses. *Science* (80- ). 352(6289):1–5.

Salvadó Z, Arroyo-López FN, Guillamón JM, Salazar G, Querol A, Barrio E. 2011. Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl Environ Microbiol.* 77(7):2292–302. doi:10.1128/AEM.01861-10.

Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, Rine J, Johnston M, Hittinger CT. 2011. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the *Saccharomyces sensu stricto* Genus. *G3* (Bethesda). 1(1):11–25. doi:10.1534/g3.111.000273.

Schluter D, Marchinko KB, Barrett RDH, Rogers SM. 2010. Natural selection and the genetics

of adaptation in threespine stickleback. *Philos Trans R Soc Lond B Biol Sci.* 365(1552):2479–86. doi:10.1098/rstb.2010.0036.

Sharon E, Chen S-AA, Khosla NM, Smith JD, Pritchard JK, Fraser HB. 2018. Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell.* 175(2):544–557.e16. doi:10.1016/j.cell.2018.08.057.

Sinha H, David L, Pascon RC, Clauder-Münster S, Krishnakumar S, Nguyen M, Shi G, Dean J, Davis RW, Oefner PJ, et al. 2008. Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics.* 180(3):1661–70. doi:10.1534/genetics.108.092932.

Smith EN, Kruglyak L. 2008. Gene-environment interaction in yeast gene expression. *PLoS Biol.* 6:810–824. doi:10.1371/journal.pbio.0060083.

Smukowski Heil C, Large CRL, Patterson K, Dunham MJ. 2018. Temperature preference biases parental genome retention during hybrid evolution. *Bioarxiv.*:1–32. doi:10.1101/429803.

Smukowski Heil CS, DeSevo CG, Pai DA, Tucker CM, Hoang ML, Dunham MJ. 2017. Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. *Mol Biol Evol.* 34(7):1596–1612. doi:10.1093/molbev/msx098.

Solieri L, Antunez O, Perez-Ortín JE, Barrio E, Giudici P. 2008. Mitochondrial inheritance and fermentative: oxidative balance in hybrids between *Saccharomyces cerevisiae* and *Saccharomyces uvarum*. *Yeast.* 25:485–500. doi:10.1002/yea.

Spirek M, Polakova S, Jatzova K, Sulo P. 2015. Post-zygotic sterility and cytonuclear

compatibility limits in *S. cerevisiae* xenomitochondrial cybrids. *Front Genet.* 5(January):1–15.  
doi:10.3389/fgene.2014.00454.

Steele DF, Butler CA, Fox TD. 1996. Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proc Natl Acad Sci U S A.* 93(11):5253–7. doi:10.1073/pnas.93.11.5253.

Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW. 2002. Dissecting the architecture of a quantitative trait locus in yeast. *Nature.* 416(6878):326–30.  
doi:10.1038/416326a.

Stern D, Orgogozo V. 2009. Is genetic evolution predictable? *Science (80- ).* 323(February):746–751.

Stern DL. 2000. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution (N Y).* 54(4):1079–1091. doi:10.1111/j.0014-3820.2000.tb00544.x.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution (N Y).* 62(9):2155–77. doi:10.1111/j.1558-5646.2008.00450.x.

Storici F, Durham CL, Gordenin DA, Resnick MA. 2003. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci.* 100(25):14994–14999. doi:10.1073/pnas.2036296100.

Stuckey S, Mukherjee K, Storici F. 2011. In vivo site-specific mutagenesis and gene collage using the delitto perfetto system in yeast *Saccharomyces cerevisiae*. Tsubouchi H, editor. 745(2):173–191. doi:10.1007/978-1-61779-129-1.

Sucena E, Stern DL. 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of *ovo/shaven-baby*. *Proc Natl Acad Sci U S A*. 97(9):4530–4.

Sulo P, Szabóová D, Bielik P, Poláková S, Šoltys K, Jatzová K, Szemes T. 2017. The evolutionary history of *Saccharomyces* species inferred from completed mitochondrial genomes and revision in the ‘yeast mitochondrial genetic code.’ *DNA Res*. 24(6):571–583. doi:10.1093/dnares/dsx026.

Suomi F, Menger KE, Monteuuis G, Naumann U, Kursu VAS, Shvetsova A, Kastaniotis AJ. 2014. Expression and Evolution of the Non-Canonically Translated Yeast Mitochondrial Acetyl-CoA Carboxylase Hfa1p. *PLoS One*. 9(12):e114738. doi:10.1371/journal.pone.0114738.

Swain Lenz D, Riles L, Fay JC. 2014. Heterochronic meiotic misexpression in an interspecific yeast hybrid. *Mol Biol Evol*. 31(6):1333–1342. doi:10.1093/molbev/msu098.

Swamy KBS, Chu W-Y, Wang C-Y, Tsai H-K, Wang D. 2011. Evidence of association between nucleosome occupancy and the evolution of transcription factor binding sites in yeast. *BMC Evol Biol*. 11(1):150. doi:10.1186/1471-2148-11-150.

Swan TM, Watson K. 1998. Stress tolerance in a yeast sterol auxotroph: Role of ergosterol, heat shock proteins and trehalose. *FEMS Microbiol Lett*. 169(1):191–197. doi:10.1016/S0378-1097(98)00483-2.

Swan TM, Watson K. 1999. Stress tolerance in a yeast lipid mutant: membrane lipids influence tolerance to heat and ethanol independently of heat shock proteins and trehalose. *Can J Microbiol*. 45(6):472–479. doi:10.1139/w99-033.

Tamura K, Subramanian S, Kumar S. 2004. Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation Clocks. *Mol Biol Evol.* 21(1):36–44.

doi:10.1093/molbev/msg236.

Thiele DJ. 1988. ACE1 regulates expression of the *Saccharomyces cerevisiae* metallothionein gene. *Mol Cell Biol.* 8(7):2745–52. doi:10.1128/MCB.8.7.2745.

Tirosh I, Barkai N. 2008. Evolution of gene sequence and gene expression are not correlated in yeast. *Trends Genet.* 24(3):109–113. doi:10.1016/j.tig.2007.12.002.

Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science* (80- ). 324(5927):659–662.

doi:10.1126/science.1169766.

Tirosh I, Sigal N, Barkai N. 2010. Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. *Mol Syst Biol.* 6(365):365.

doi:10.1038/msb.2010.20.

Tirosh I, Weinberger A, Bezalel D, Kaganovich M, Barkai N. 2008. On the relation between promoter divergence and gene expression evolution. *Mol Syst Biol.* 4(159):159.

doi:10.1038/msb4100198.

Turner RB, Smith DL, Zawrotny ME, Summers MF, Posewitz MC, Winge DR. 1998. Solution structure of a zinc domain conserved in yeast copper-regulated transcription factors. *Nat structural Biol.* 5(7):551–555.

Verghese J, Abrams J, Wang Y, Morano KA. 2012. Biology of the heat shock response and

protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. *Microbiol Mol Biol Rev.* 76(2):115–58. doi:10.1128/MMBR.05018-11.

Vigh L, Maresca B, Harwood JL. 1998. Does the membrane's physical state control the expression of heat shock and other genes? *Trends Biochem Sci.* 23(10):369–374. doi:10.1016/S0968-0004(98)01279-1.

De Virgilio C, Bürckert N, Bell W, Jenö P, Boller T, Wiemken A. 1993. Disruption of TPS2, the gene encoding the 100-kDa subunit of the trehalose-6-phosphate synthase/phosphatase complex in *Saccharomyces cerevisiae*, causes accumulation of trehalose-6-phosphate and loss of trehalose-6-phosphate phosphatase activity. *Eur J Biochem.* 212(2):315–23. doi:10.1111/j.1432-1033.1993.tb17664.x.

Vögtle F-N, Burkhart JM, Gonczarowska-Jorge H, Kücükköse C, Taskin AA, Kopczynski D, Ahrends R, Mossmann D, Sickmann A, Zahedi RP, et al. 2017. Landscape of submitochondrial protein distribution. *Nat Commun.* 8(1):290. doi:10.1038/s41467-017-00359-0.

Vorachek-Warren MK, McCusker JH. 2004. DsdA (D-serine deaminase): a new heterologous MX cassette for gene disruption and selection in *Saccharomyces cerevisiae*. *Yeast.* 21(2):163–71. doi:10.1002/yea.1074.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Andersson M, An P, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 420(6915):520–562. doi:10.1038/nature01262.

Weiss C V., Roop JI, Hackley RK, Chuong JN, Grigoriev I V, Arkin AP, Skerker JM, Brem RB. 2018. Genetic dissection of interspecific differences in yeast thermotolerance. *Nat Genet.*



doi:10.1038/s41588-018-0243-4.

Welch J, Fogel S, Buchman C, Karin M. 1989. The CUP2 gene product regulates the expression of the CUP1 gene, coding for yeast metallothionein. *EMBO J.* 8(1):255–60.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: Raw material for evolution. *Mol Ecol.* 15(5):1197–1211. doi:10.1111/j.1365-294X.2006.02868.x.

Williams EH, Butler CA, Bonnefoy N, Fox TD. 2007. Translation initiation in *Saccharomyces cerevisiae* mitochondria: Functional interactions among mitochondrial ribosomal protein Rsm28p, initiation factor 2, methionyl-tRNA-formyltransferase and novel protein Rmd9p. *Genetics.* 175(3):1117–1126. doi:10.1534/genetics.106.064576.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature.* 430(6995):85–8. doi:10.1038/nature02698.

Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet.* 13(1):59–69. doi:10.1038/nrg3095.

Wittkopp PJ, Vaccaro K, Carroll SB. 2002. Evolution of yellow Gene Regulation and Pigmentation in *Drosophila*. *Curr Biol.* 12(18):1547–1556. doi:10.1016/S0960-9822(02)01113-2.

Wolters JF, Charron G, Gaspary A, Landry CR, Fiumera AC, Fiumera HL. 2018. Mitochondrial Recombination Reveals Mito–Mito Epistasis in Yeast. *Genetics.* 209(1):307–319. doi:10.1534/genetics.117.300660.

Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in

*Saccharomyces cerevisiae*. BMC Genomics. 16(1):451. doi:10.1186/s12864-015-1664-4.

Wright D. 2015. The genetic architecture of domestication in animals. Bioinform Biol Insights. 9:11–20. doi:10.4137/BBI.S28902.

Wu B, Buljic A, Hao W. 2015. Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. Mol Biol Evol. 32(10):2559–2570. doi:10.1093/molbev/msv127.

Xi L, Fondufe-Mittendorf Y, Xia L, Flatow J, Widom J, Wang JP. 2010. Predicting nucleosome positioning using a duration Hidden Markov Model. BMC Bioinformatics. 11:346. doi:10.1186/1471-2105-11-346.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591. doi:10.1093/molbev/msm088.

Yun Y, Adesanya TMA, Mitra RD. 2012. A systematic study of gene expression variation at single nucleotide resolution reveals widespread regulatory roles for uAUGs. Genome Res. 22(6):1089–1097. doi:10.1101/gr.117366.110.

Zeevi D, Lubliner S, Lotan-Pompan M, Hodis E, Vesterman R, Weinberger A, Segal E. 2014. Molecular dissection of the genetic mechanisms that underlie expression conservation in orthologous yeast ribosomal promoters. Genome Res. 24(12):1991–9. doi:10.1101/gr.179259.114.

Zeng Z, Liu J, Stam L, Kao C, Mercer JM, Laurie CC. 2000. Genetic architecture of a morphological shape difference between two *Drosophila* species. Genetics. 154(January):299–

310.

Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. 2011. Regulatory Variation Within and Between Species. *Annu Rev Genomics Hum Genet.* 12(1):327–346. doi:10.1146/annurev-genom-082908-150139.

Zhou X, Peris D, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. 2016. in silico Whole Genome Sequencer and Analyzer (iWGS): a Computational Pipeline to Guide the Design and Analysis of de novo Genome Sequencing Studies. *G3.* 6(November):3655–3662. doi:10.1534/g3.116.034249.

Zubko EI, Zubko MK. 2014. Deficiencies in mitochondrial DNA compromise the survival of yeast cells at critically high temperatures. *Microbiol Res.* 169(2–3):185–95. doi:10.1016/j.micres.2013.06.011.

# Appendix

## Supplementary materials for Chapter 2

Table S2-1. GO terms enriched in temperature-responsive genes.

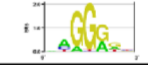

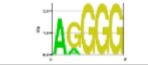




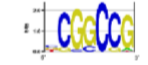
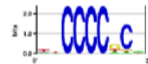

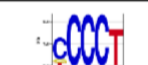



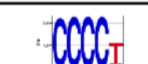

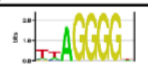

Gene set	GOID	GO_term	Cluster frequency	Background frequency	P-value	FDR	Genes annotated to the term
22°C> 37°C	6553	lysine metabolic process	5 out of 110 genes, 4.5%	9 out of 5055 background genes, 0.2%	0.00026	0	<i>LYS2, LYS20, LYS12, LYS1, LYS9</i>
22°C> 37°C	9085	lysine biosynthetic process	5 out of 110 genes, 4.5%	9 out of 5055 background genes, 0.2%	0.00026	0	<i>LYS2, LYS20, LYS12, LYS1, LYS9</i>
22°C> 37°C	1901605	alpha-amino acid metabolic process	12 out of 110 genes, 10.9%	101 out of 5055 background genes, 2.0%	0.00071	0	<i>LYS2, CHA1, CIT2, LYS20, ARO10, ARG4, LYS12, LYS1, GAD1, CIT1, LYS9, IDH2</i>
22°C> 37°C	9066	aspartate family amino acid metabolic process	7 out of 110 genes, 6.4%	36 out of 5055 background genes, 0.7%	0.00463	0	<i>LYS2, CHA1, LYS20, ARO10, LYS12, LYS1, LYS9</i>
22°C> 37°C	19878	lysine biosynthetic process via aminoadipic acid	4 out of 110 genes, 3.6%	8 out of 5055 background genes, 0.2%	0.0068	0	<i>LYS2, LYS20, LYS1, LYS9</i>
22°C> 37°C	6457	protein folding	10 out of 110 genes, 9.1%	88 out of 5055 background genes, 1.7%	0.00842	0	<i>AHA1, MDJ1, BTN2, SSA2, HSP104, HSC82, SIS1, STI1, HSP82, CUR1</i>

37°C> 22°C	19030 46	meiotic cell cycle process	19 out of 101 genes, 18.8%	244 out of 5055 background genes, 4.8%	9.93E- 05	0	<i>DTR1, SPO23, SPR28, MEI4, DMC1, RCK1, CLB1, SPO13, SPS100, SPO16, HOP1, GAS2, SMA2, CSM3, REC114, FKS3, NDJ1, MPC54, YOR338W</i>
37°C> 22°C	51321	meiotic cell cycle	19 out of 101 genes, 18.8%	249 out of 5055 background genes, 4.9%	0.00014	0	<i>DTR1, SPO23, SPR28, MEI4, DMC1, RCK1, CLB1, SPO13, SPS100, SPO16, HOP1, GAS2, SMA2, CSM3, REC114, FKS3, NDJ1, MPC54, YOR338W</i>
37°C> 22°C	44702	single organism reproductive process	20 out of 101 genes, 19.8%	283 out of 5055 background genes, 5.6%	0.00022	0	<i>DTR1, SPO23, SPR28, MEI4, DMC1, RCK1, CLB1, GPA1, SPO13, SPS100, SPO16, HOP1, GAS2, SMA2, CSM3, REC114, FKS3, NDJ1, MPC54, YOR338W</i>
37°C> 22°C	9062	fatty acid catabolic process	5 out of 101 genes, 5.0%	15 out of 5055 background genes, 0.3%	0.00324	0	<i>ADR1, POX1, POT1, CIT3, ICL2</i>


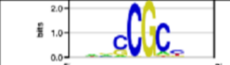




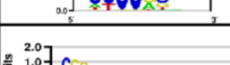


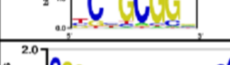
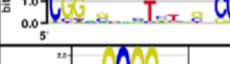


37°C> 22°C	44281	small molecule metabolic process	28 out of 101 genes, 27.7%	598 out of 5055 background genes, 11.8%	0.00399	0	<i>ADE1, GAL10, VID24, GCV1, ADR1, GPP2, ARG5,6, SER3, AGX1, POX1, GND2, POT1, ARG3, SIP4, MET5, MET14, MTD1, PCK1, PCD1, PIG1, FBP1, PHO84, GCV2, CAT8, BIO4, ARG1, CIT3, ICL2</i>
---------------	-------	---	----------------------------------	--	---------	---	---

Results (GO terms, p-values and FDR) are from the Saccharomyces Genome Database using genes significantly up or down at 22°C compared to 37°C compared to a background set of all genes with expression data.

Figure S2-1.

Class	Motif	TF	Logo	$p_1$	$p_2$	$R^2$	Predict ASE?	$P_{ASE}$	$R^2_{ASE}$
22°C > 37°C	YMR037C_1380	MSN2		7.70E-23	0.00172	0.0289	No		
22°C > 37°C	YPR086W_1327	SUA7		5.00E-08	0.00320	0.0272	No		
22°C > 37°C	YPL230W_509	USV1		1.07E-26	0.00334	0.0270	No		
22°C > 37°C	YPR052C_879	NHP6A		0.00107	0.00377	0.0267	No		
22°C > 37°C	YGL073W_411	HSF1		0.00803	0.00460	0.0261	No		
22°C > 37°C	YDR096W_562	GIS1		1.19E-26	0.00952	0.0240	No		
22°C > 37°C	YAL051W_2060	OAF1		3.34E-11	0.0108	0.0236	No		
22°C > 37°C	YCR106W_506	RDS1		2.44E-10	0.0174	0.0223	No		
22°C > 37°C	YJR127C_575	RSF2		1.92E-08	0.0182	0.0221	No		
22°C > 37°C	YDR043C_2148	NRG1		1.09E-16	0.0198	0.0219	No		
22°C > 37°C	YMR182C_531	RGM1		3.62E-23	0.0215	0.0216	No		
22°C > 37°C	YDR216W_576	ADR1		7.21E-09	0.0217	0.0216	No		
22°C > 37°C	YDR034C_865	LYS14		2.32E-05	0.0237	0.0213	No		
22°C > 37°C	YDL170W_486	UGA3		5.23E-06	0.0252	0.0211	No		
22°C > 37°C	YKL062W_518	MSN4		5.66E-23	0.0268	0.0210	No		
22°C > 37°C	YLL054C_816	YLL054C		3.61E-09	0.0315	0.0205	No		
22°C > 37°C	YER169W_547	RPH1		3.33E-19	0.0319	0.0204	No		
22°C > 37°C and show allele by temperature interaction	YDR421W_725	ARO80		0.00459	0.0495	0.177	Yes	0.00278	0.114

(Figure S2-1 continued)

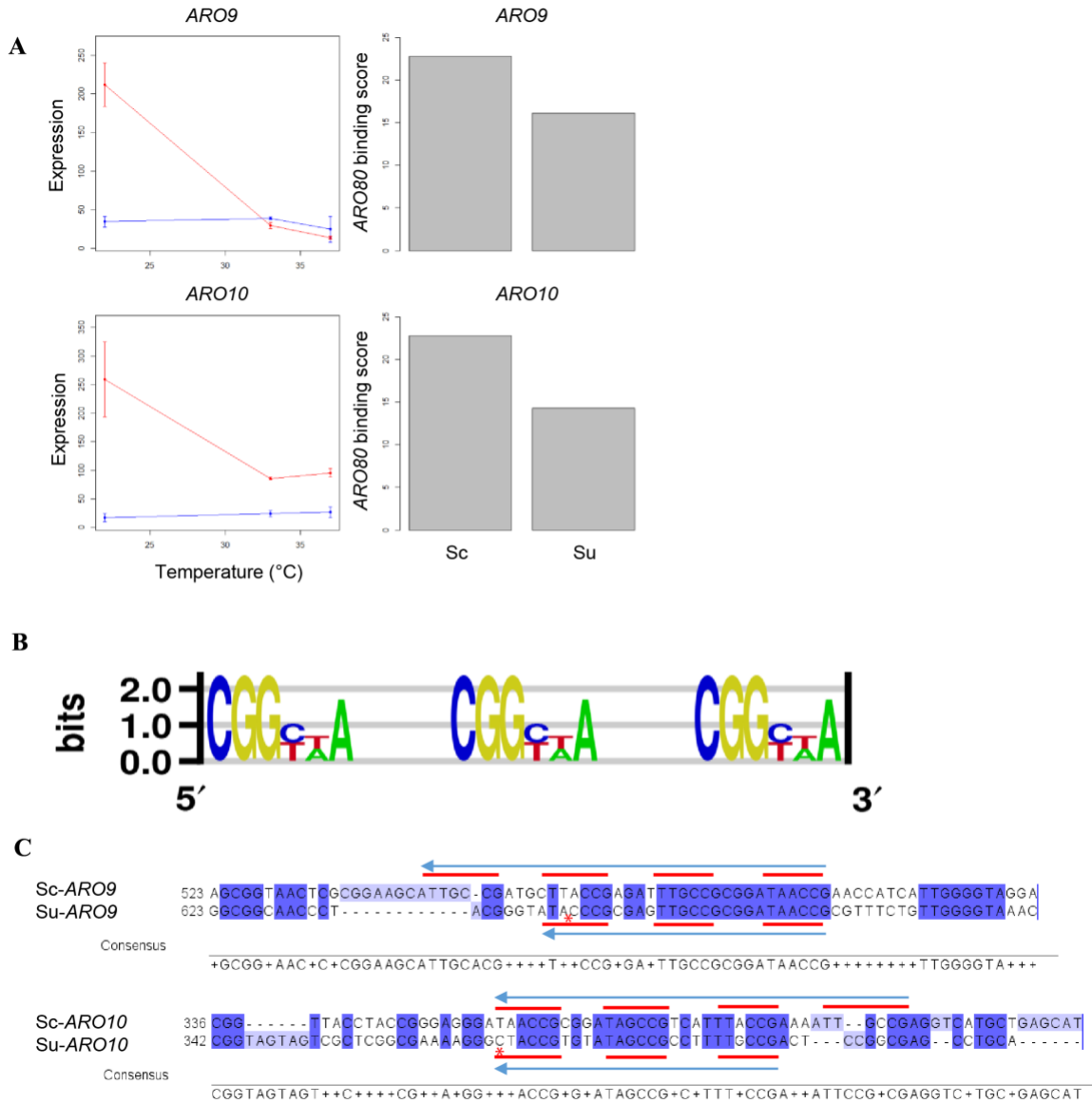
Class	Motif	TF	Logo	p <sub>1</sub>	p <sub>2</sub>	R <sup>2</sup>	Predict ASE?	p <sub>ASE</sub>	R <sup>2</sup> <sub>ASE</sub>
22°C < 37°C	YDR216W_576	ADR1		0.00669	0.000296	0.0359	No		
22°C < 37°C	YDR207C_2239	UME6		0.00131	0.00081	0.0328	No		
22°C < 37°C	YGL035C_2142	MIG1		0.0134	0.00327	0.0286	Yes	0.0347	0.0181
22°C < 37°C	YGR067C_2191	YGR067C		0.0361	0.00690	0.0264	Yes	0.00842	0.0247
22°C < 37°C	YDR463W_660	STP1		1.66E-05	0.00758	0.0261	No		
22°C < 37°C	YGL209W_2143	MIG2		0.00359	0.00858	0.0257	Yes	0.00420	0.0282
22°C < 37°C	YAL051W_2060	OAF1		0.0328	0.0110	0.0249	No		
22°C < 37°C	YER028C_2144	MIG3		0.0153	0.0110	0.0249	Yes	0.000678	0.0366
22°C < 37°C	YHR178W_2068	STB5		0.00551	0.0254	0.0223	No		
22°C < 37°C	YPL248C_1510	GAL4		0.00184	0.0260	0.0223	No		
22°C < 37°C	YDL170W_486	UGA3		0.00330	0.0266	0.0222	No		
22°C < 37°C	YDL170W_651	UGA3		0.00372	0.0321	0.0216	No		
22°C < 37°C	YML081W_2194	TDA9		0.0248	0.0333	0.0215	Yes	0.0348	0.0177

**Figure S2-1. Transcription factor binding motifs associated with temperature effects.** p<sub>1</sub>, Holm–Bonferroni corrected p-values of Mann–Whitney tests for enrichment in the given temperature-responsive class. p<sub>2</sub>, Holm–Bonferroni corrected p-values of linear regression between binding scores and temperature effects. R<sup>2</sup>, adjusted R-squares of linear regression between binding scores and temperature effects. p<sub>ASE</sub>, Holm–Bonferroni corrected p-values of linear regression between binding divergence and allele effects. R<sup>2</sup><sub>ASE</sub>, adjusted R-squares of



linear regression between binding divergence and allele effects. Motif names and logos were downloaded from YeTFaSCo database (De Boer and Hughes 2012). TF, transcription factor. ASE, allele-specific expression.

Figure S2-2.



**Figure S2-2. Correlated binding site and expression changes for *ARO9* and *ARO10*.** **A.** Expression and *ARO80* binding site scores for *ARO9* and *ARO10*. Expression is represented by normalized read counts for the *S. cerevisiae* allele (red) and *S. uvarum* allele (blue). Barplots show *ARO80* binding scores for *S. cerevisiae* (Sc) and *S. uvarum* (Su). **B.** *ARO80* motif. **C.** Sequence alignment of *ARO80* binding sites, from *S. cerevisiae* (Sc) and *S. uvarum* (Su). For

both *ARO9* and *ARO10*, the lower binding scores in *S. uvarum* are caused by single nucleotide changes (red asterisks). The *S. cerevisiae* binding sites are tetrameric and the *S. uvarum* sites are trimeric (red lines).

## Supplementary materials for Chapter 3

### Supplementary Text

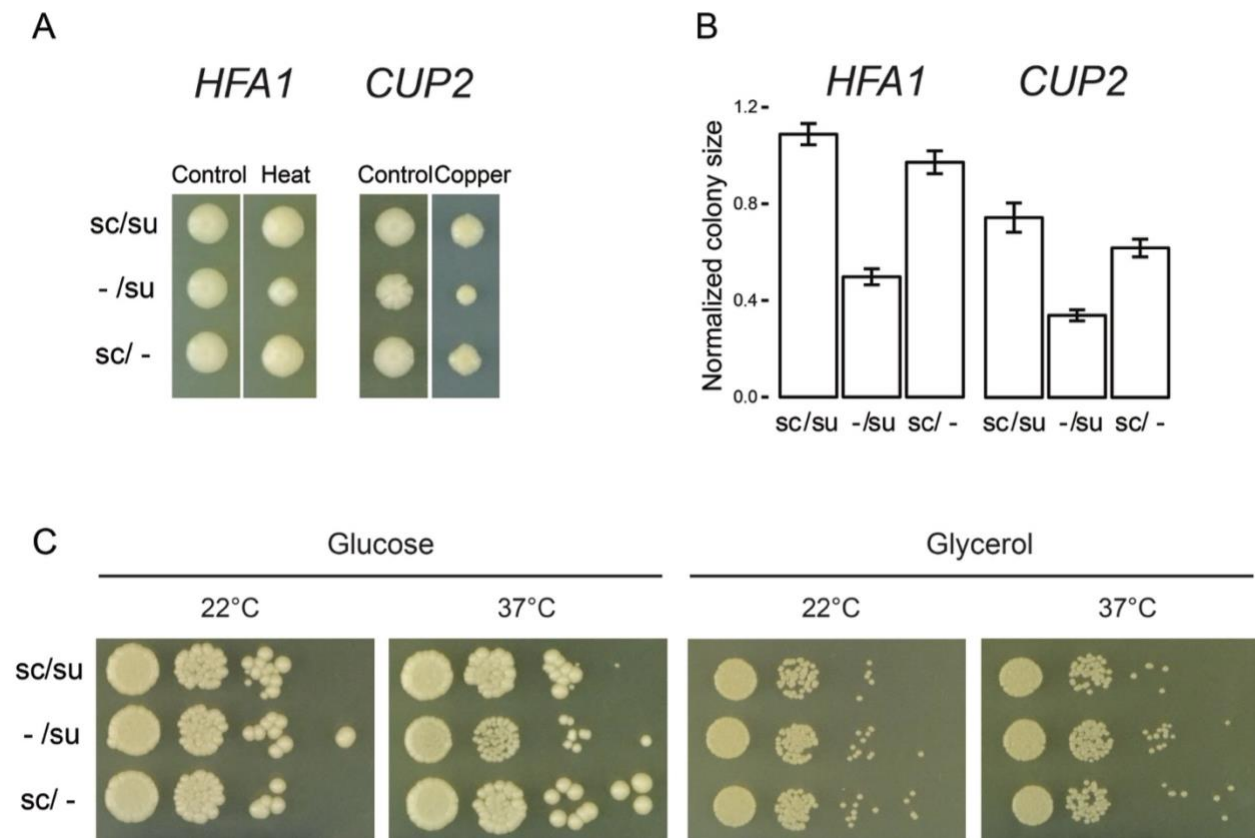
#### High petite rate of *S. uvarum* mitotype and its association with *ORF1*

*Saccharomyces* yeast strains generate petites spontaneously at a rate of ~1%, and variants in nuclear genes can affect petite rates (Dimitrov et al. 2009). We observed an extremely high petite rate in the hybrid with *S. uvarum* mitotype (48-61%, sometimes >90%), while the hybrid carrying *S. cerevisiae* mtDNA rarely generates petites (Fig. S3-5A). The high petite rate associated with *S. uvarum* mtDNA is only seen in the interspecific hybrid, but not pure strains *S. uvarum*, suggesting a dominant incompatibility in mtDNA inheritance between hybrid nuclear genomes and *S. uvarum* mtDNA. However, we were able to isolate a few *S. cerevisiae* and *S. uvarum* hybrids that carried mostly *S. uvarum* mitochondrial genes but did not exhibit a high petite rate. These strains arose at a frequency of 1%, so they are likely spontaneous recombinants. Whole genome sequencing showed that they all carry *S. cerevisiae ORF1*, but the rest of their mitochondrial genome is *S. uvarum* (Fig. S3-5C). This result suggests a strong link between *S. cerevisiae ORF1* and mtDNA inheritance. In the 90 recombinants generated from mutant crosses, we also observed a strong correlation between *S. cerevisiae ORF1* and low petite rates, although there were exceptions (Fig. S3-5B).

The possible inheritance phenotype adds to our understanding of the interesting biology of *ORF1*. *ORF1* (F-*SceIII*) was suggested to encode a free-standing homing endonuclease (Bordonné et al. 1988). The best-known homing endonuclease is I-*SceI* ( $\omega$ ), which promotes its spread to homing-less mitochondrial genomes (Jacquier and Dujon 1985). *ORF1* (F-*SceIII*) has been proposed to mediate mitochondrial recombination based on the high frequency of

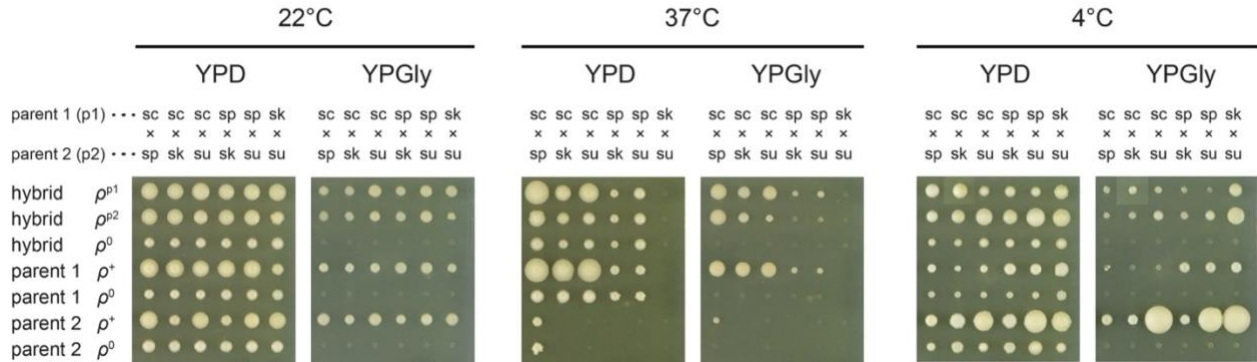
interspecific mitochondrial recombinants at the start of *ORF1* in wild *Saccharomyces* species and in a synthetic hybrid of *S. cerevisiae* × *S. mikatae* (Peris, Arias, et al. 2017; Peris, Moriarty, et al. 2017). Although further work will be needed to demonstrate that *ORF1* affects mitochondrial inheritance, this activity would imply co-evolution between a selfish element and its host (Burt and Koufopanou 2004).

Figure S3-1.



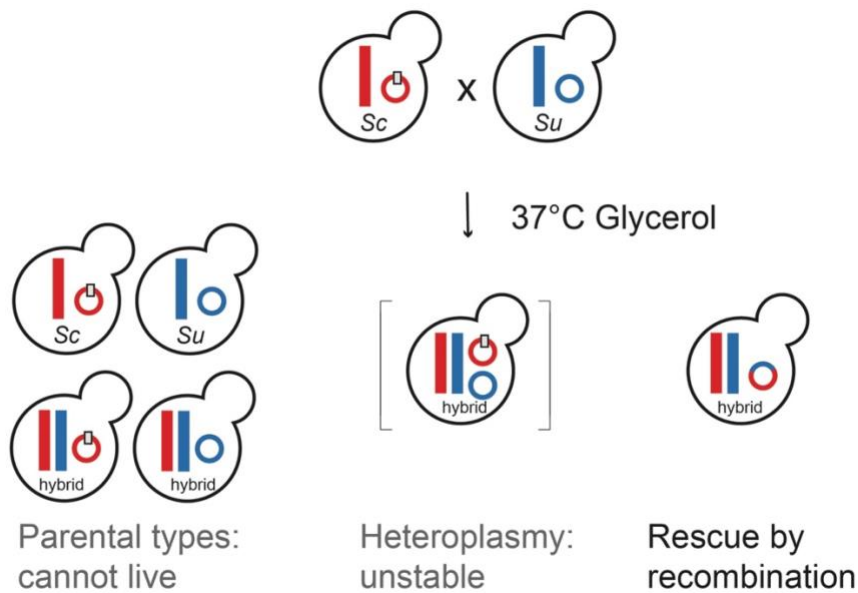
**Figure S3-1. Reciprocal hemizygosity test of *HFA1* and *CUP2*.** (A) Hemizygotes with either the *S. cerevisiae* allele (sc/-) or *S. uvarum* allele (-/su) and a wild-type hybrid (sc/su) were compared under the same conditions as the non-complementation screen. Growth is after 5 days. (B) Heat or copper resistance was measured by colonies sizes normalized to control condition (22°C YPD), with error bars representing the standard deviation of 6 biological replicates. (C) *HFA1* hemizygotes differed in heat sensitivity on glucose but not glycerol medium. Cells were plated at 1:10 dilution. Growth is after 3 days.

Figure S3-2.



**Figure S3-2. Fermentative and respiratory growth of interspecific hybrids with reciprocal mitotypes at different temperatures.** Interspecific hybrids between *S. cerevisiae* (sc), *S. paradoxus* (sp), *S. kudriavzevii* (sk), and *S. uvarum* (su) with either parental mitotype ( $\rho^{p1}$  or  $\rho^{p2}$ ) or no mtDNA ( $\rho^0$ ) were grown on YPD and YPGly plates for 5 days (22°C and 37°C) or 124 days (4°C). Growth of parent species and their petites are shown for comparison. The 4°C images of *S. cerevisiae* × *S. kudriavzevii* hybrid with *S. cerevisiae* mtDNA (sc × sk  $\rho^{p1}$ ) were replaced with images from a biological replicate plated in the same configuration because the original colony was contaminated.

Figure S3-3.

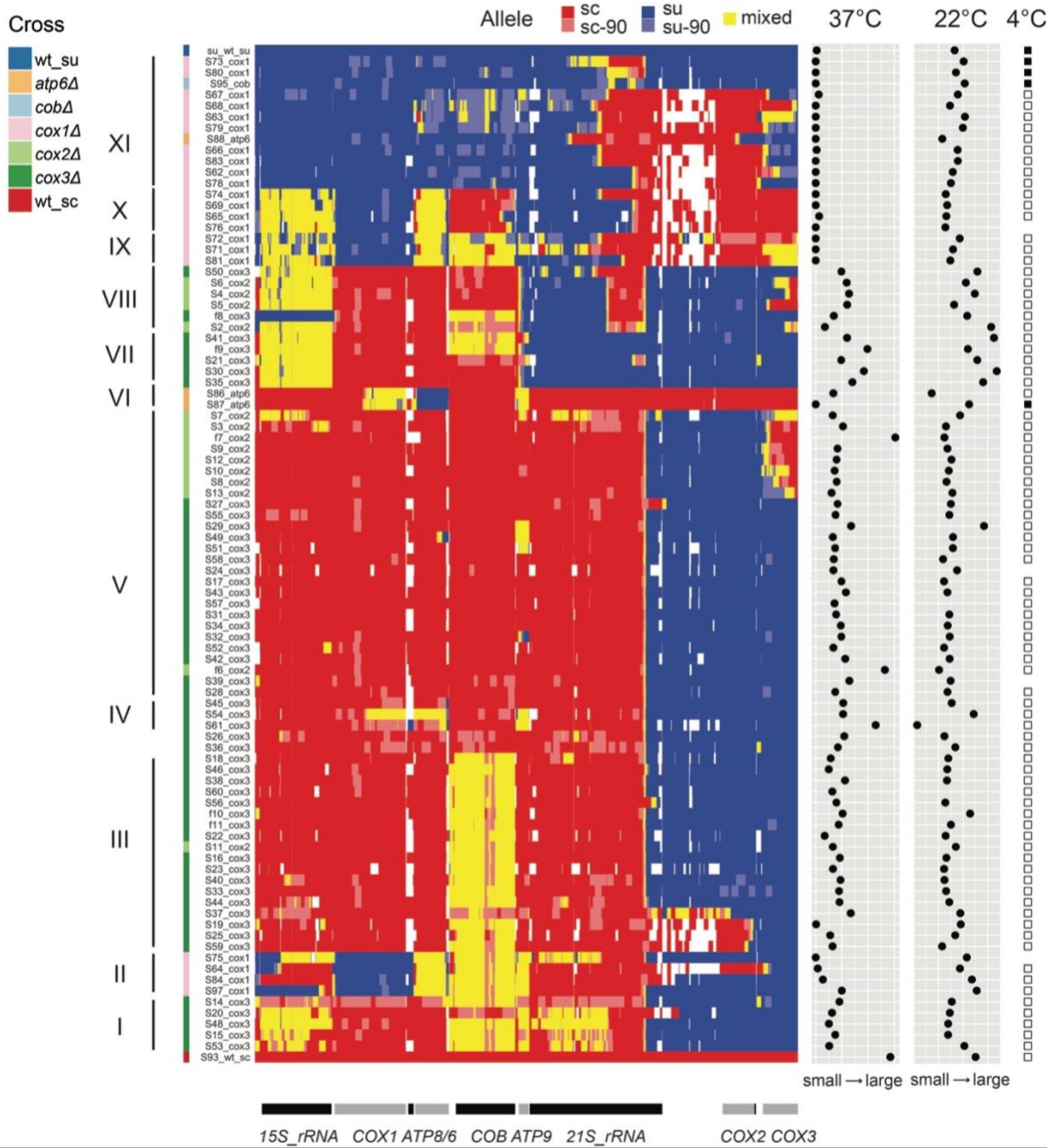


**Figure S3-3. Rescue of *S. cerevisiae* (sc) mitochondrial knockouts by recombination with *S. uvarum* (su) mitotypes.** Upon crossing *S. cerevisiae* with *S. uvarum*, hybrids have unstable heteroplasmy; parental types do not grow at 37°C on glycerol, but recombinants can rescue the *S. cerevisiae* deficiency and the *S. uvarum* temperature sensitivity.



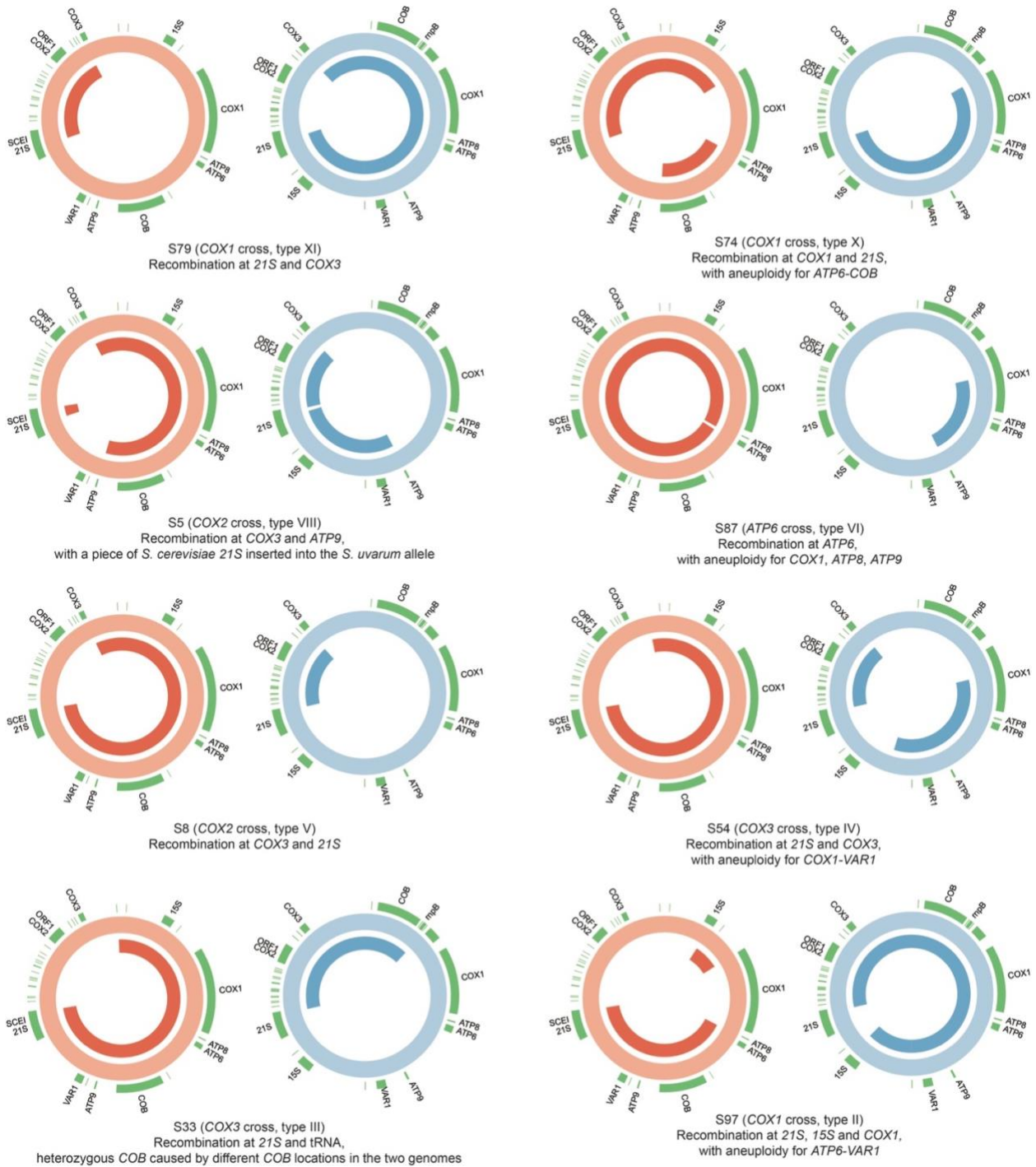
Figure S3-4.

A



(Figure S3-4 continued)

B

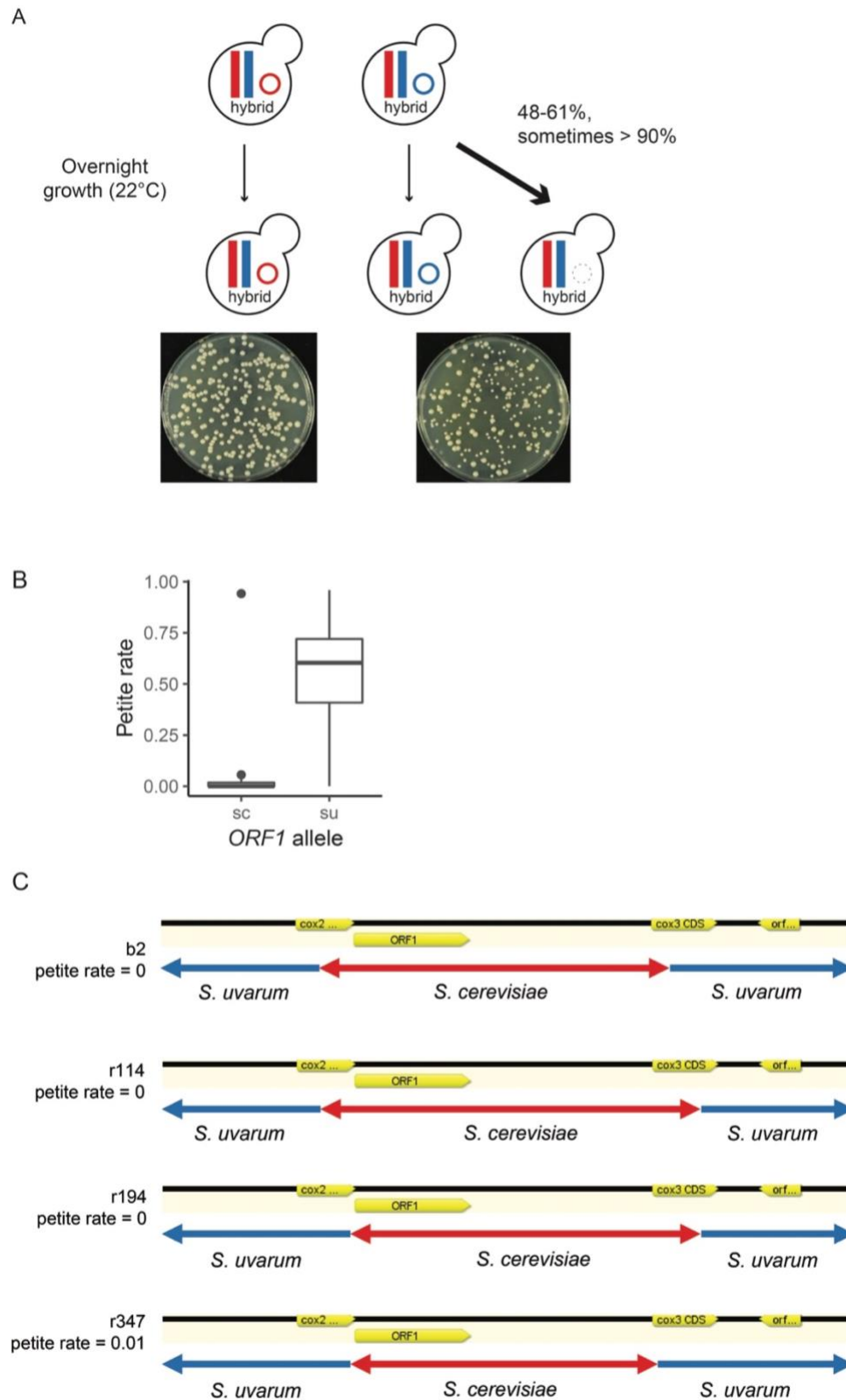


**Figure S3-4. Recombinant genotypes and examples of recombination breakpoints. (A)**

Recombinants were manually classified into 11 genotype groups and breakpoints for 8

representatives were identified by manual inspection. Strains were labeled by the trials (“f” for initial trial and “S” for second trial) and mutant crosses in which they were generated. Phenotype panels are shown as in Fig. 2B, with the addition of 22°C colony sizes. **(B)** Representative recombinant genomes are shown. Outer circles represent the reference mitochondrial genomes (red for *S. cerevisiae*, blue for *S. uvarum*) and inner circles show coverage of a given recombinant. Note *15S rRNA* and *COB* are at different positions in the two reference genomes.

Figure S3-5.

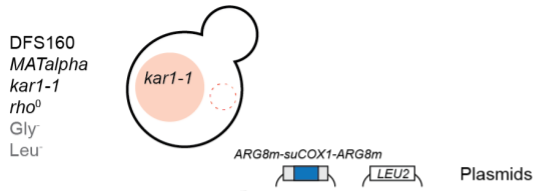


**Figure S3-5. High petite rate of *S. uvarum* mitotype and its association with *ORF1*. (A)**

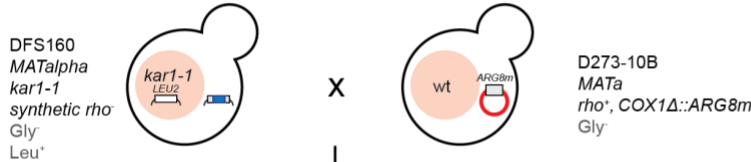
Petite rate in a 22°C overnight culture is high for the hybrid with a *S. uvarum* mitotype (blue circle), while the hybrid with a *S. cerevisiae* mitotype (red circle) rarely generates petites (dotted circle). (B) Petite rates associate with *ORF1* alleles in 90 recombinants generated by knockout crosses. sc, *S. cerevisiae*; su, *S. uvarum*. (C) Four spontaneous recombinants carrying *S. cerevisiae ORF1* showed low petite rates; the rest of their mitochondrial genome is *S. uvarum*.

Figure S3-6.

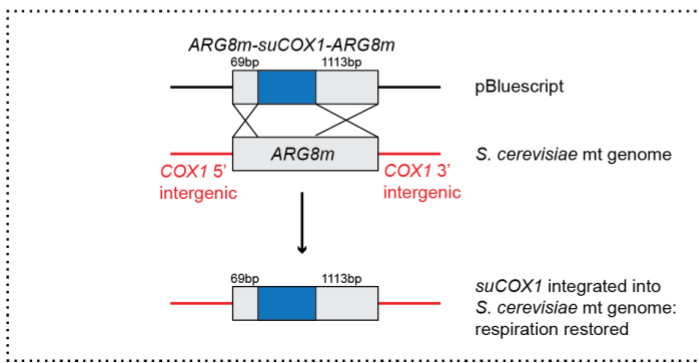
A



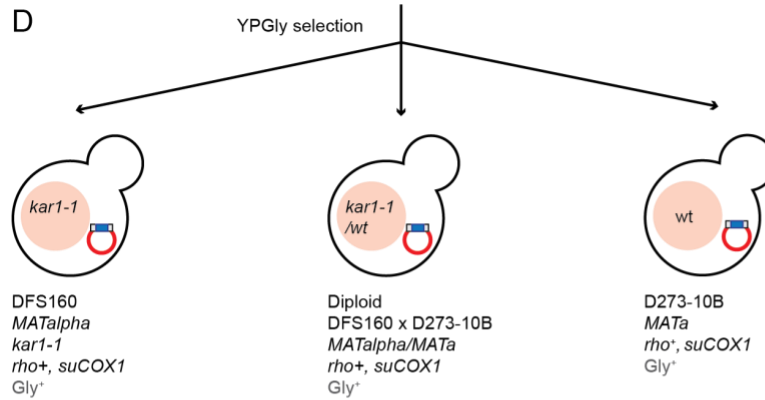
B



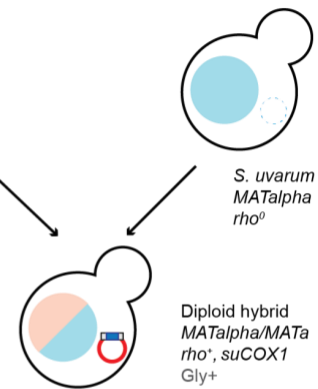
C



D



E

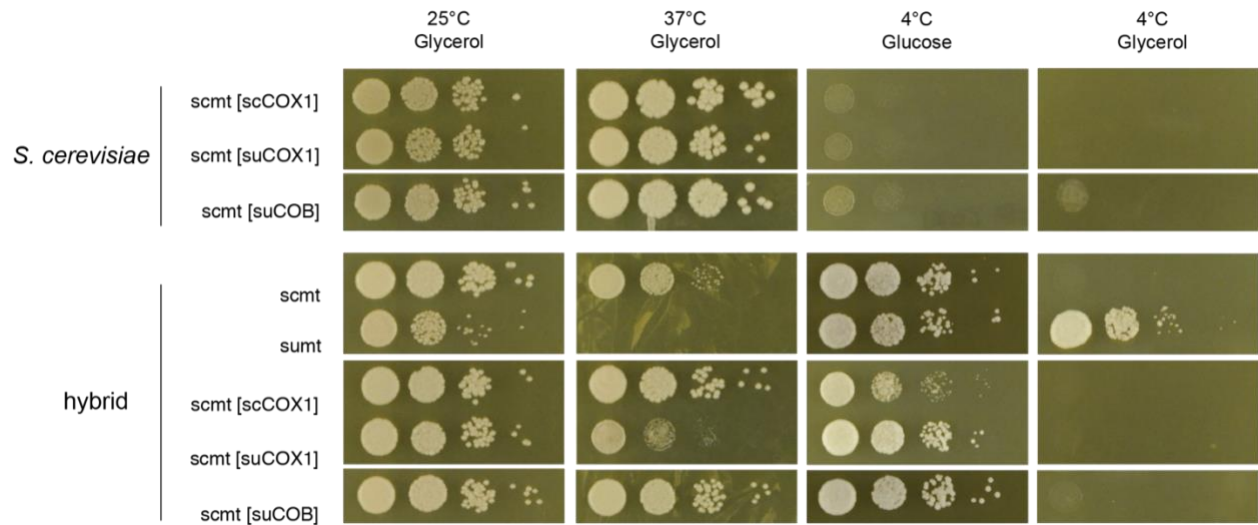


Phenotypic assays:

25°C YPD  
37°C YPGly  
4°C YPGly

**Figure S3-6. Procedure for mitochondrial allele replacement.** (A) Biolistic transformation of the mitochondrial construct with a *LEU2* plasmid. (B) Leu<sup>+</sup> colonies were mated to *S. cerevisiae* mitochondrial knockouts. (C) The allele of interest was integrated into the mitochondrial genome via homologous recombination. (D) Integrated alleles were selected by rescue of respiration. (E) *MATa* mitochondrial genome transformants were crossed to *S. uvarum*.

Figure S3-7.



**Figure S3-7. Background-dependent allele effects of *COX1*.** *S. cerevisiae* diploids and hybrids carrying allele replacements and two wild-type controls were plated with 1:10 serial dilution and incubated at indicated temperatures. Growth is after 4 days for 25°C and 37°C, 25 days for 4°C on glucose, and 53 days for 4°C on glycerol. sc, *S. cerevisiae*; su, *S. uvarum*; mt, mtDNA. Alleles in the brackets were integrated into their endogenous loci in *S. cerevisiae* mtDNA.



Table S3-1. Aneuploidy in the recombinants.

<b>Strain</b>	<b>Increased 37°C growth compared to similar genotypes?</b>	<b>Cross</b>	<b>Duplicated chromosome</b>	<b>Mitochondrial interacting genes carried on the chromosome<sup>1</sup></b>	<b>Reference</b>
<b>S29</b>	Yes	<i>cox3</i> Δ	<i>S. cerevisiae</i> chrIX	<i>MRS1</i>	(Chou et al. 2010)
<b>S53</b>	No	<i>cox3</i> Δ	<i>S. cerevisiae</i> chrV	<i>MRX1</i>	(Jhuang et al. 2017)
<b>S54</b>	No	<i>cox3</i> Δ	<i>S. cerevisiae</i> chrIX	<i>MRS1</i>	(Chou et al. 2010)
<b>S61</b>	Yes	<i>cox3</i> Δ	<i>S. cerevisiae</i> chrIX	<i>MRS1</i>	(Chou et al. 2010)
<b>S97</b>	Yes	<i>cox1</i> Δ	<i>S. uvarum</i> chr10	<i>PET309</i>	(Jhuang et al. 2017)

<sup>1</sup> Only genes with known incompatibilities were listed.

Table S3-2. Strains used in this study.

Species	Strain	Background	Nuclear genotype	Mitochondrial genotype	Reference
<i>S. cerevisiae</i>	YJF173	S288C	<i>MATa ho- ura3-52</i>	<i>rho</i> <sup>+</sup>	(Engle and Fay 2012)
<i>S. cerevisiae</i>	YJF153	YPS163 <sup>1</sup>	<i>MATa hoΔ::dsdAMX4</i>	<i>rho</i> <sup>+</sup>	(Swain Lenz et al. 2014)
<i>S. cerevisiae</i>	PJD1 <sup>2</sup>	D273-10B	<i>MATa ade2-101 ade3-24 leu2-3,112 ura3-52</i>	<i>rho</i> <sup>+</sup>	(Williams et al. 2007)
<i>S. cerevisiae</i>	XPM10b <sup>2</sup>	D273-10B	<i>MATa arg8::hisG leu2-3,112 lys2 ura3-52</i>	<i>rho</i> <sup>+</sup> , <i>cox1Δ::ARG8m</i>	(Perez-Martinez et al. 2003)
<i>S. cerevisiae</i>	HMD22 <sup>2</sup>	D273-10B	<i>MATa arg8::hisG his3-deltaHinDIII leu2-3,112 lys2 ura3-52</i>	<i>rho</i> <sup>+</sup> , <i>cox2Δ::ARG8m</i>	(Bonney and Fox 2000)
<i>S. cerevisiae</i>	DFS168 <sup>2</sup>	D273-10B	<i>MATa arg8Δ::URA3 his4-519 leu2-3,112 ura3Δ</i>	<i>rho</i> <sup>+</sup> , <i>cox3Δ::ARG8m</i>	(Steele et al. 1996)
<i>S. cerevisiae</i>	CAB59 <sup>2</sup>	D273-10B	<i>MATa arg8::hisG leu2-3,112 lys2 ura3-52</i>	<i>rho</i> <sup>+</sup> , <i>cobΔ::ARG8m COX1 minus at least one intron</i>	(Ding et al. 2009)
<i>S. cerevisiae</i>	MR6 <sup>3</sup>	W303-1B for nuclear; FY1679 for mitochondrial	<i>MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 CAN1 arg8::HIS3</i>	<i>rho</i> <sup>+</sup>	(Rak et al. 2007)
<i>S. cerevisiae</i>	MR6 ΔATP6 <sup>3</sup>	W303-1B for nuclear; FY1679 for mitochondrial	<i>MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 CAN1 arg8::HIS3</i>	<i>rho</i> <sup>+</sup> , <i>atp6Δ::ARG8m</i>	(Rak et al. 2007)
<i>S. cerevisiae</i>	MR6 ΔATP8 <sup>3</sup>	W303-1B for nuclear; FY1679 for mitochondrial	<i>MATa ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 CAN1 arg8::HIS3</i>	<i>rho</i> <sup>+</sup> , <i>atp8Δ::ARG8m</i>	(Rak and Tzagoloff 2009)
<i>S. cerevisiae</i>	DFS160	DBY947	<i>MATa ade2-101 kar1-1 leu2Δ ura3-52 arg8Δ::URA3</i>	<i>rho</i> <sup>0</sup>	(Steele et al. 1996)
<i>S. uvarum</i>	YJF1449	CBS7001	<i>MATa hoΔ::NatMX</i>	<i>rho</i> <sup>+</sup>	(Scannell et al. 2011)

<i>S. uvarum</i>	YJF1450	CBS7001	<i>MATa</i> <i>hoΔ::NatMX</i>	<i>rho</i> <sup>+</sup>	(Scannell et al. 2011)
<i>S. uvarum</i>	YJF2600	CBS7001	<i>MATa</i> <i>hoΔ::NatMX</i> <i>trp1Δ::hphMX4</i>	<i>rho</i> <sup>+</sup>	This study
<i>S. uvarum</i>	YJF2601	CBS7001	<i>MATa</i> <i>hoΔ::NatMX</i> <i>trp1Δ::hphMX4</i>	<i>rho</i> <sup>+</sup>	This study
<i>S. uvarum</i>	YJF2760	CBS7001	<i>MATa</i> <i>hoΔ::NatMX</i> <i>trp1Δ::hphMX4</i>	<i>rho</i> <sup>0</sup>	This study
<i>S. uvarum</i>	YJF2785	BMV58 <sup>4</sup>	<i>MATa</i> <i>hoΔ::dsdAMX4</i>	<i>rho</i> <sup>+</sup>	This study
<i>S. paradoxus</i>	YJF694	N17	<i>MATa hoΔ::lox-KAN-lox</i>	<i>rho</i> <sup>+</sup>	This study
<i>S. paradoxus</i>	YJF695	N17	<i>MATa hoΔ::lox-KAN-lox</i>	<i>rho</i> <sup>+</sup>	This study
<i>S. kudriavzevii</i>	YJF2779	CR85 <sup>5</sup>	<i>MATa</i> <i>hoΔ::hygMX4</i> <i>ura3</i>	<i>rho</i> <sup>+</sup>	(Lopes et al. 2010)

<sup>1</sup>YPS163 is an oak tree isolate provided by P. Sniegowski.

<sup>2</sup>provided by Tom D. Fox.

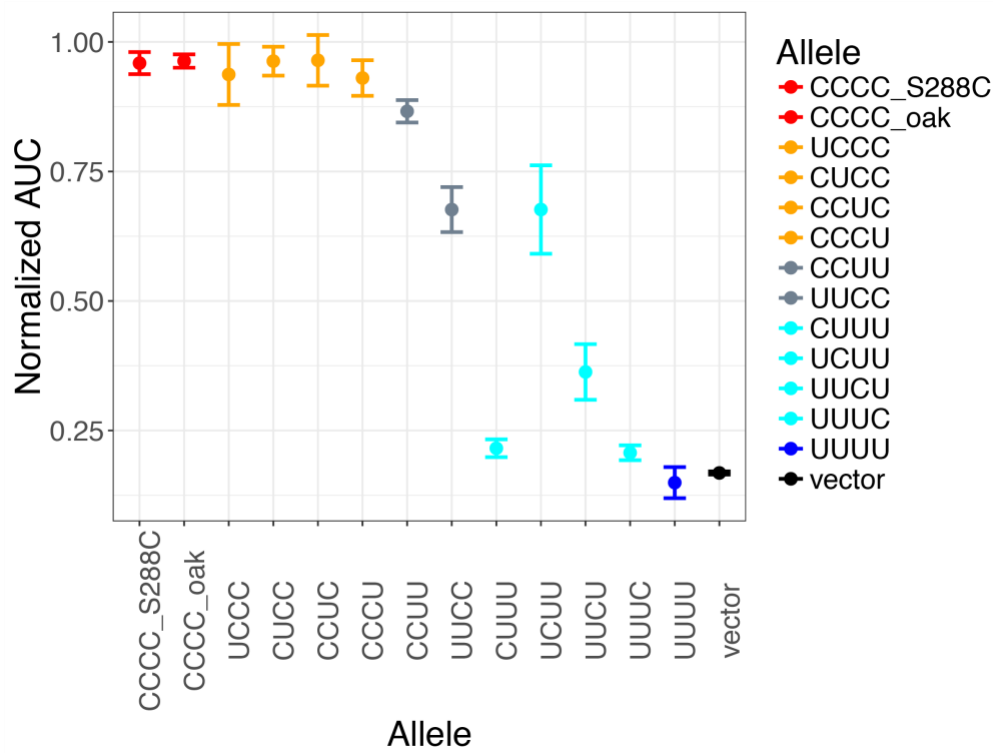
<sup>3</sup>provided by Alexander Tzagoloff.

<sup>4</sup>BMV58 is a wine isolate provided by Javier Alonso del Real Arias and Amparo Querol; used in constructing *S. paradoxus* × *S. uvarum* hybrids.

<sup>5</sup>CR85 is a Spanish *Quercus ilex* bark isolate provided by Amparo Querol.

## Supplementary materials for Chapter 4

Figure S4-1.



**Figure S4-1. Copper resistance of chimeric constructs in *S. cerevisiae*.** Copper resistance of the chimeras was examined in an *S. cerevisiae* *CUP2* knockout strain in 0.5 mM copper sulfate. The resistance was measured by normalized area under curve (AUC), with points representing the mean of three biological replicates and error bars representing 95% confidence interval. The colors are based on the number of *S. cerevisiae* segments in the chimeras (red = 4, orange = 3, grey = 2, light blue = 1, blue or black = 0).

## Supplementary materials for Chapter 5

Verified sequence for *S. uvarum* *ura3Δ::sc-ura3-140* locus. Red is the truncated *S. cerevisiae* *URA3* ORF. Unshaded is *S. cerevisiae* sequence; shaded is downstream *S. uvarum* flanking sequence. Upstream *S. uvarum* flanking sequence is not shown because it was not covered by Sanger reads.

```
TATTTATGGTGAAGAATAAGTTTTGACCATCAAAGAAGGTTAATGTGGCTGTGGTTTTCAGGGTC
CATAAAGCTTTTTCAATTCATCTTTTTTTTTTTTATTCTTTTTTTTGATTCCGGTTTCCTTGAAAT
TTTTTTGATTCGGTAATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGG
TATATATACGCATATGTAGTGTGAAGAAACATGAAATTGCCAGTATTCTTAACCCAAGTCA
CAGAACAAAAACCTGCAGGAAACGAAGATAAATCATGTCGAAAGCTACATATAAGGAACGTGCT
GCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAACT
TGTGTGCTTCATTGGATGTTTCGTACCACCAAGGAATTAATTGAGTGA TCGTAACGTAGCGTAGA
CTAACGTGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTT
CGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCTGCGGGTGTATAC
AGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCAGGTATTGTTAGCG
GTTTGAAGCAGGCGGCAGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATT
GTCATGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGC
GACAAAGATTTTTGTTATCGGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACG
ATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCAACAGTA
TAGAACCGTGATGATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGAAGAGGACTATTT
GCAAAGGGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATT
TGAGAAGATGCGGCCAGCAAACTAAAAAAGTATTATAAGTAAATGCATGTATACTAAACTC
ACAACCTTAGAGCTTCAATTTAATTATATCAGTTATTACCCGGAATCTCGGTCGTAATGATTTT
TATAATGACGAAAAAAAAAAAAATTGAAAAGAAAAAGCTTCATGGCCTTTATAAAAAGGAACCAT
CCAATACCTCGCCAGAACCAAGGAAATTATCCAATATCATAGCAAAATTTAAAGACATCCAAA
CGGCTCAGACCAAACAGAACAAGACAGGACCGCAGAGATGGACGCACTAAATCTAAAGAACAA
CAAGAGTTCAGAAAAGTGGTGGAAACAAAAGCAAATGAAGGATTTTCATGCGTTTATACTCCAAC
TAGTAGAAAGGTGTTTACGGACTGTGTCAACGATTTCACAACTTCCAAGCTAACCAACAAGGA
GCAAACATGCATCATGAAGTGTTTCAGAGAAGTTCCTCAAGCACAGCGAGCGTGTTCGGACAACGT
TTCCAAGAACAGAACGCCGCTTTGGGCCAAGGCCTGGGGCGTTAATATGTACTGGCGTATATAA
CTATATATCTAATCGGGTCTGTTTCTAGCATGTAAATATAAAAACAAATAAATCAATAGTATT
ATTACCT
```