The genetic dissection of trait differences between species of *Saccharomyces* yeasts

By

Carly V Weiss

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Rachel B. Brem, co-chair
Professor John W. Taylor, co-chair
Professor Adam P. Arkin
Professor Benjamin Blackman

Fall 2018

Abstract

The genetic dissection of trait differences between species of *Saccharomyces* yeasts

by

Carly V Weiss

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Rachel B. Brem, co-chair
Professor John W. Taylor, co-chair

Unbiased genetic dissection of widely observable phenotypic traits in the wild has long been the goal of evolutionary geneticists. Mice, bacteria, rice and stickleback fishes have historically been among the prime model organisms in the field. Here, we leverage the recent surge of *Saccharomyces* yeast as a model genus in ecology and evolution to begin answering questions about the genetic basis of ancient trait differences that have evolved between species, over long evolutionary time scales. Specifically, we want to know what genetic mechanisms evolution has used to create new traits in the distant past, and what biological functions have been the focus of adaptation in the past. I introduce the field and our questions of interest in the introductory Chapter 1. Next, Chapter 2 delves deeper into the methods that have been used in the past to dissect interspecies genetics, reviewing the literature and drawing general conclusions from what we have learned so far. In Chapter 3, we develop a new application of the reciprocal hemizygote test on a genome-wide scale to drill down to the single gene level and dissect the ability of *S. cerevisiae* to grow at high temperatures relative to other *Saccharomyces* species. We uncover a suite of housekeeping genes genetically responsible for this derived phenotype and reveal a likely defect in cell division in *S. paradoxus* as the culprit for cell death at high temperatures. Finally, in Chapter 4, we investigate two more species-specific differences: 1) resistance to the drug benomyl, a microtubule poison and 2) cold tolerance. Unexpectedly, while microtubules are the molecular target of benomyl, it is the genes encoding water channels in the cell membrane tha seem to be at the genetic root of these phenotypes. Our data begin to connect the dots between the growth advantage in *S. paradoxus* in benomyl and at low temperatures, relative to *S. cerevisiae*.

Acknowledgements

Although including an Acknowledgements section in this dissertation is technically not required, I think I would be doing a great disservice not to write about the people who helped me along the way.

First, I must thank my thesis advisor, Rachel Brem. When I think back on the last years of my life doing this thesis work, I could not have asked for a better mentor. Suffice it to say that she is not only the hardest working person I have ever known, but also her love of science and drive to produce quality work has served as a source of inspiration to me during the hardest moments of my Ph.D. I knew I would be working with some of the best minds in the world when I came to Berkeley, but what I didn't realize was that I would also get to work alongside someone who is just genuinely such a great human being, and for that I will be forever thankful. I know that if I can channel a fraction of what Rachel has tried to teach me, I will be successful in science. All I can say is thank you, so very much, for everything you have done for me.

Next, I would like to acknowledge the other lab members who contributed their time and energy to this work. Jeremy Roop, a previous graduate student in the lab, created the foundation upon which most of this work was based. His work allowed me to take off immediately upon joining the lab and accomplish a lot in a short period of time. Rylee Hackley and Julie Chuong are two extremely talented technicians who have worked alongside me over the course of these projects, tirelessly helping with the most tedious of tasks that would have otherwise driven me insane. Ryan Kenneally, Jake Kim, Justin Hong, Claire Dubin, Anna Flury and Faisal AlZaben are undergraduates who have worked with me over the last several years. Being able to act as a mentor to each of them helped me develop my skills as a manager, and they all helped bring a sense of community to our otherwise small lab. Melanie Abrams, Tyler Hilsabeck, Arjun Sasikumar and Chris Nelson are other lab members who provided great conversation and helpful feedback along the way. Lastly, Jeffrey Skerker has always made himself available to help with technical details, and having another high-level researcher to get advice from has been great, scientifically and professionally.

Finally, my family and friends have always been so supportive, even when they had no clue what I am working on or why. My sister, Ashley, has been a pillar of support and a place to rant about anything and everything. My parents put me through undergraduate and supported me moving to Berkeley from Toronto, even though it meant they would see me much less often. To my friends – you know who you are, and I know I would have quit long ago without you and all the happiness and meaning you have brought to my life outside of the lab.

Chapter 1

An introduction to yeast evolutionary biology in the context of interspecific phenotypic divergence

Why do we want to understand the genetic basis of natural phenotypic variation?

A wondrous and wide variety of life forms exist on the Earth, from archaea that can grow at 113°C[1] and cheetahs able to run up to 75 miles per hour[2], to more than 300,000 species of beetle[3], all of which have come to be from the humble beginnings of the first atoms of the Big Bang. The field of evolutionary biology addresses how such extreme variation in the traits (ie. observable characteristics) of living organisms has come to exist throughout the history of the planet and the sub-field of evolutionary genetics works to characterize the molecular basis of how organisms adapt to their environment. Although Darwin didn't know it when he wrote On the Origin of Species in 1859, evolution by natural selection works largely through deoxyribonucleic acid (DNA), the heritable material within each living organism. Modern evolutionary genetics specifically interrogates the alterations at the DNA level that have evolved to generate a specific trait, with the goal of finding the link between an organism's DNA sequence (genotype) and its observable characteristics (phenotype). Geneticists refer to this as mapping genotype to phenotype.

Thus far, the field has made great strides in mapping the genetic underpinnings of observable traits and using this knowledge for novel innovations in biomedicine, industry, and agriculture. For example, detection in a patient of the E4 allele of the APOE gene, associated with Alzheimer's disease, can now be used to support a positive diagnosis of the disease[4]. In cases where a trait is governed by variation at a single location in the genome ("locus"), it can be relatively easy to understand the genetic basis of that trait, and the literature abounds with these types of studies. However, this simple paradigm of "one-trait, one-gene" has been slowly reworked over the years, with it now being generally accepted that most varying traits are governed by multiple different loci, in complex combinations, sometimes numbering in the 100's or 1000's[5] In reality, very few examples exist of a complex trait being fully or even partially understood at the genetic level.

In addition, the vast majority of work in the field has focused on differences that occur between individuals of a given species, which by definition must have evolved relatively recently. This focus, although while well-intentioned for the sake of simplicity, is limiting, as some of the most interesting trait differences exist only between reproductively isolated populations, i.e. between species instead of within a single species. This includes traits that may have evolved long ago, in the common ancestor of many species. For example, as they diversified, different subspecies of rice (*Oryza sativa*) acquired distinct resistance to bacterial infection[6], and species of the pathogen Cryptococcus specialized to different ranges of human hosts[7]. A greater understanding of the genetic determinants of traits differing between species would not only allow us to

engineer novel and beneficial technologies, such as disease treatments or pathogen-resistant crops, it would also allow us to peer further back into evolutionary time than is currently possible and better understand how complex traits are built by evolution.

Saccharomyces yeasts as a model for ecology and evolution

Saccharomyces cerevisiae has been long studied as a eukaryotic model organism in molecular genetics and cell biology. In recent years the *Saccharomyces* genus has also emerged as a new model in ecology and evolution[8]. More and more strains are being collected from diverse environments all over the world, with a recent paper publishing genomic and phenotypic data on over 1,000 unique isolates from Taiwan to the Netherlands and everywhere in between[9]. Combined with the depth of knowledge and tools that have accrued from *S. cerevisiae*'s history as a model organism in cell biology and molecular genetics, it has become possible to answer questions about basic evolutionary principles using this system that cannot be answered using more complex organisms or less well-studied microbes.

Specifically, we decided to study species within the *sensu stricto* clade of *Saccharomyces*, focusing on the well-known *S. cerevisiae* as well as its sister species, *S. paradoxus.* There were several reasons for this decision. The first is that the two species are separated by 5-10 million years of evolution[10], giving us a good reason to expect that there will be many phenotypic differences between species. Another reason is that their genomes are mostly syntenic, allowing us to use chromosome position in the genome along with nucleotide sequence identify homologous genes between the two species. In addition, both species are easy to work with and manipulate in the lab, have many previously validated genetic tools, and have a rich history in the literature. Protocols already existing for *S. cerevisiae* for DNA transformation, gDNA isolation and culturing are easy to adapt to *S. paradoxus* with small changes. And although the number of collected isolates around the globe does not rival *S. cerevisiae*, several groups have started compiling *S. paradoxus* isolates from diverse geographic locations. Finally, we predicted that by choosing the species closest in evolutionary age to *S. cerevisiae* as a comparison, we could minimize the chances that any hybrid strains formed between the two would be genetically abnormal (i.e. form aneuploidies or be genetically unstable); the relevance of this choice will become apparent in the next section.

What is reciprocal-hemizygosity analysis?

Reciprocal-hemizygosity analysis was first conceived and applied by Steinmetz *et. al* in 2002[11]. In this seminal paper, the authors determined the genetic basis of the high-

growth phenotype at high temperatures of a clinical isolate of *S. cerevisiae* relative to a lab strain. In general, reciprocal-hemizygosity analysis (also called the "reciprocal hemizygote test") is a way to compare the phenotypic effects of allelic differences between two individuals (Figure 1). Put another way, if two individuals have differing DNA sequences at a given locus, the reciprocal hemizygote test is a way to determine the effects of those differing sequences on a phenotype of interest. To do this at a given locus, the two individuals, strains or species must be mated to form a hybrid, half of whose genome comes from each of the respective parents. In this background, two strains are generated, each containing an interrupted or deleted copy of each parent's allele of the given locus. These strains are hemizygous since they remain diploid everywhere in the genome except at the locus of interest, where they are considered haploid, and are referred to as reciprocal since each lacks only one parent's allele, with its remaining allele the other parent. By comparing the phenotypes of these two reciprocal hemizygote strains, one can conclude whether DNA sequence variants at the manipulated locus contribute to the trait of interest, since variants at the locus are the only genetic difference between the two reciprocal hemizygote strains. In this way, it is possible to link genetic differences between species to a phenotypic difference between them in a clean and well-controlled experimental system.

There are a few caveats to reciprocal-hemizygosity analysis that are worth mentioning in the context of applying this technique to studying interspecific trait variation. Firstly, the two target species must be able to be mated to form a viable hybrid. Therefore, it is important to assess the genetic distance between the two species under consideration and understand any incompatibilities that may arise in a hybrid (e.g. heterosis). Another requirement is that it must be possible to genetically manipulate the hybrid in order to interrupt a locus of interest. Finally, it is important to remember that any genes mapped through reciprocal-hemizygosity analysis are inherently mapped in a hybrid genetic background, and any hits must be validated in the original parent genetic backgrounds. Epistasis (genetic interaction) can have effects on phenotypic differences observed in the hybrid background such that they do not necessarily recapitulate in the purebred parents. Even in light of these caveats, reciprocal hemizygosity analysis is still a remarkably robust method for interrogating the phenotypic consequences of genetic differences. Unfortunately, it remains a relatively obscure method and has only been used on a candidate gene basis to test the potential involvement of certain suspected loci in trait variation[12,13]. The main advancement of the work described in this dissertation is to apply this method on a genome-wide scale in an unbiased manner, screening the genome for genes potentially involved in an ancient trait.

The contents of this thesis

The following chapters detail my attempts to contribute to the topics I have discussed above. Chapter 2 is a literature review on the various approaches existing in the field to dissect interspecies differences, their relative pros and cons, and the fundamental principles of evolution we have learned so far from each. Chapter 3 describes my main project, where we discover part of the genetic basis of a thermotolerance difference between *S. cerevisiae* and *S. paradoxus.* The genes we found were not in expected heat stress response pathways, and instead were essential genes important for basic housekeeping functions in the cell. Chapter 4 describes the current progress of a newer project, aimed at discovering the genetic basis of a difference in resistance to a microtubule stressor (the fungicide benomyl) between *S. cerevisiae* and *S. paradoxus.* The story emerging from our latest results suggests that the loss of benomyl tolerance in *S. cerevisiae* may be connected mechanistically to this species' parallel loss of cold tolerance. I detail what future efforts are ongoing on this project.

In conclusion, I have been driven over the last several years to answer basic questions about evolution. What molecular mechanisms have been used in the distant past to create the traits we see in nature in the present? Does a complex trait typically evolve through many genes or a single gene? Are the changes more likely to be within protein-coding portions of genes, or within those segments of the genome that regulate their expression? What biological processes in the cell can be adapted to form new traits? Our results add to the growing consensus that complex traits are built using a variety of loci throughout the genome. We show that each locus can have a variable contribution to the trait, and that these effects are indeed significantly dependent on the genomic background in which the variant is located. Although I have barely scratched the surface in the following work, I trust that this humble contribution has moved the field forward, if only by a small amount.

Figures

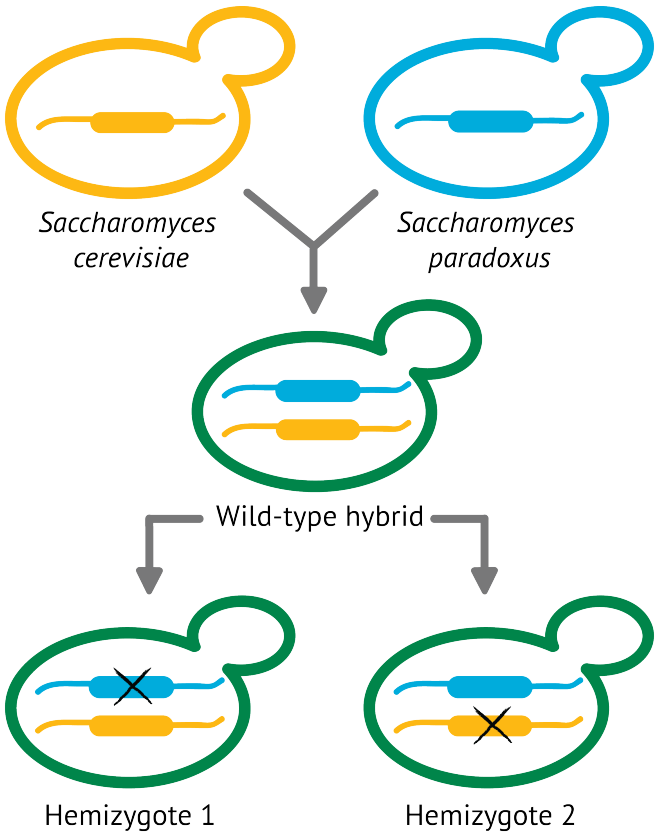


Figure 1. Reciprocal hemizygotes are created through the generation of an F1 hybrid between two parent species. Two parent species, here shown as *S. cerevisiae* and *S. paradoxus*, in yellow and blue respectively, are mated to form an F1 hybrid, in green. The wild-type hybrid contains a single copy of each of the parent's genomes. At a given locus, interrupting a single copy from one of the two parents (black X) creates a hemizygous strain ('1', interruption in the *S. paradoxus* allele, leaving the *S. cerevisiae* allele intact), diploid at the rest of the genome except for the locus of interest. Creating the reciprocal strain ('2', interruption in the *S. cerevisiae* allele, leaving the *S. paradoxus* allele intact) with an interruption in the same locus, in the other parent's allele, generates a pair of reciprocal hemizygotes. By comparing the phenotypes of the two strains, one can interrogate the phenotype effects of allelic variation at the interrupted locus.

References

1 Blöchl, E. *et al*. *Pyrolobus fumarii,* gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1(1)**, 14-21 (1997).

2 Sharp N. Timed running speed of a cheetah (*Acinonyx jubatus*). *J Zool Lond* **241**, 493–494 (1997).

3 Stork, N.E. *et al*. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proc Natl Acad Sci USA* **112(24)**, 7519–7523 (2015).

4 Corder, E.H. *et al*. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923 (1993).

5 Price, A. L, Spencer, C. C. A. & Donnelly, P. Progress and promise in understanding the genetic basis of common diseases. *Proc R Soc B* **282** doi: 10.1098/rspb.2015.1684 (2015).

6 Cao, Y. et al. The Expression Pattern of a Rice Disease Resistance Gene Xa3/Xa26 Is Differentially Regulated by the Genetic Backgrounds and Developmental Stages That Influence Its Function. *Genetics* **177(1)**, 523-533 (2007).

7 Kwon-Chung, K.J. et al. *Cryptococcus neoformans* and *Cryptococcus gattii*, the Etiologic Agents of Cryptococcosis. *Cold Spring Harb Perspect Med* **4(7)**, a019760 (2014).

8 Replanksy, T., Koufopanou, V., Greig, D. & Bell, G. *Saccharomyces sensu stricto* as a model system for evolution and ecology. *Trends Eco Evo* **23(9),** 494-501 (2008).

9 Peter, J., et. al. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature **556**, 339-344 (2018).

10 Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Bethesda)* **1**, 11-25, doi:10.1534/g3.111.000273 (2011).

11 Steinmetz, L. M., Sinha, H., Richards, D. R., Spiegelman, J. I., Oefner, P. J., McCuskter, J. H. & Davis, R. W. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416,** 326-330 (2002).

12 Nadeau, N. J. & Jiggins, C. D. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet* **26**, 484-492, doi:10.1016/j.tig.2010.08.004 (2010).

13 Wray, G. A. Genomics and the Evolution of Phenotypic Traits. *Annual Review of Ecology, Evolution, and Systematics* **44**, 51-72 (2013).

Chapter 2

A review of the current literature on methods to dissect interspecific trait variation

Why study interspecies variation?

For decades, geneticists have sought to understand how the phenotypes of living organisms are encoded in their DNA. Powerful genomic screening approaches, as well as classical hypothesis-driven cell and molecular biology, have revealed genes *required* for many traits, especially in model species. But in the vast majority of cases, we don't know how the genetic determinants that were *sufficient* for the trait when it arose—that is, how the trait evolved in the first place, and how it has been refined over time.

The field of evolutionary genetics has as its goal to understand the machinery with which evolution builds adaptive traits, *i.e.* their genetic architecture. Has an adaptation arisen owing to DNA sequence differences at a single place in the genome, or many? How much of a new beneficial trait can be attributed to changes in the sequence or structure of a protein, versus changes in the quantity or location of that protein in the cell? How much of the effect of any given genomic variant observed depends on changes that have occurred in parallel elsewhere in the genome?

In fact, questions of exactly this kind have been addressed quite deeply in the study of standing variation within populations. These advances have come thanks to modern linkage and association mapping methods, which screen the genome for variants co-inherited with a trait of interest, across recombinant progeny from matings between individuals of a given species. A given such trait is often not an adaptation at all, but rather is the product of neutral or slightly deleterious alleles that arose relatively recently and are segregating by genetic drift in large populations. In some cases, however, individuals have undergone a bona fide adaptation in a new environment so recently that they can still mate with relatives that have maintained the ancestral program (a situation referred to as local adaptation). In this scenario the adaptive alleles can be found by a genome-wide linkage scan in crosses between the derived and ancestral populations. The latter have revealed substantial insights into the genetic mechanisms of adaptation, and several excellent reviews have been written on the subject[1-4].

Importantly, by construction, this local adaptation literature centers on evolutionarily "young" traits within populations. As has become clear from long-term evolution experiments in *Escherichia coli*, it can take some adaptations a very long time to evolve – on the order of tens of thousands of generations in one example of citrate metabolism – due to the necessity for the causal mutations to accumulate in a specific order[5]. For a complete picture of the molecular mechanisms of evolution in the wild, we likely need to complement what we know about recent adaptations with the study of traits that have been tweaked and refined over millions of years. But to date, doing genetics over long timescales has posed a key challenge in the field. Long-diverged populations often become reproductively isolated from one another, in which case we call them distinct species. Any such case is then refractory to linkage and association methods, which rely on recombination between interfertile individuals.

The search for genes underlying trait variation between species is thus at the forefront of modern evolutionary genetic research, where the potential for new emerging paradigms is among the greatest, and yet experiments are the hardest. In this review, we cover the advances that have been made in the mapping of genotype to phenotype between species, structuring our survey in terms of the main experimental approaches in the field. Since for each method the literature is sizeable, we draw on a few representative examples where the impact of allelic variation between species has been functionally validated. We use these case studies to illustrate the benefits and pitfalls of each strategy.

Candidate-gene approaches

We start our survey with the situation in which the genetic basis of interspecific trait variation can be guessed from first principles. This bears out when genes underlying a phenotype in a given species are already known, and sequence differences in any such locus between species serve as good candidates for the causal basis of phenotypic divergence.

For example, Daugherty *et. al.*[6], targeting differential viral susceptibility between mammals, drew on the extensive literature characterizing the *IFIT* family of innate immune response genes. Different mammalian species have different numbers of *IFIT* genes, each of which is induced in response to viral infection. Functionally, each is thought to inhibit viral mRNA translation (potentially by binding to viral translation machinery or binding to viral mRNA, depending on which *IFIT* gene), but it is still unknown if they mediate overall viral susceptibility differences. The authors functionally characterized IFIT1 protein variants in human and mouse with respect to their ability to bind to 2'O-methylation of mRNA caps and showed that these functional differences led to distinct antiviral specificities in a non-human primate cell line system. This provided a well-supported case study of functional variants in coding (as opposed to regulatory) regions. Yet, it remains to be seen if natural variation at *IFIT* genes actually underlies human and mouse viral susceptibilities on a whole-organism level, how much trait variation can be explained by *IFIT*, and what additional genes are involved.

In another rigorous recent example, Campesan *et. al.*[7] used as a jumping-off point a history of reports of the *nonA* gene, which encodes a putative RNA-binding protein and when knocked out perturbed courtship songs in in *D. melanogaster*. The authors hypothesized that *nonA* variants could underlie differences in courtship songs between fly species, and tested this notion by swapping the regulatory and coding regions of *nonA* from *D. virilis* into a *D. melanogaster nonA*$^{-}$ mutant. Results revealed that the transgene transferred some, but not all, song features of *D. virilis* into *D. melanogaster*. This was a satisfying validation of the importance of *nonA* to courtship song divergence, although the remaining complex genetic architecture remains unknown.

Palgrave *et. al.* focused on the contrast between domestic pigs (*Sus scrofa*), which are susceptible to African swine fever virus (ASFV), and wild warthogs, which are

robustly tolerant of the virus[8]. Previous observational studies had highlighted several genes involved in ASFV immune response in *S. scrofa*, which the authors considered good candidate determinants of the interspecific differences. One such locus, *RELA,* which encodes a subunit of the signaling complex NF-κB, harbored several non-synonymous changes between domestic pig and warthog. Working in a heterologous cell line, the authors tested the functional impact of variants of this locus by transfecting either pig or warthog RELA protein into cells and measuring both basal and induced NF-κB activity via a luciferase reporter. They found that one of these amino acid variants was sufficient for striking differential NF-κB activity, but didn't test ASFV resistance. This stands as another instance in which one locus was demonstrated to contribute to interspecific trait variation, though to date, demonstrating the relevance of the *RELA* variants *in vivo* has been out of reach.

These examples highlight the strengths and weaknesses of the candidate-gene paradigm. One obvious limitation is the bias for traits that are already well understood at the mechanistic level, and for genes whose a role in a trait is well-characterized (*e.g.* from a mutagenesis screen). Yet even with any one of the latter in hand, the candidate approach may simply fail: just because one knows a gene to be required for a trait in the organism where it's observed today, there's no guarantee that evolution used alleles at this locus to build the trait in the ancestor. Furthermore, the best ways we have to test this hypothesis first-hand—evaluating whether derived alleles at a given locus are sufficient to reconstitute a trait in a naive background—are often too laborious to be done across dozens or hundreds of candidates. Thus many success stories from the candidate-gene approach center on a single "lucky" gene choice that validates early in the process. On the plus side, a study design focused on just one or a few of these successful candidate hits can be a virtue, in that it frees up resources to dissect at the subgenic level exactly which variants are causal for the trait of interest. As such, the molecular genetics of these candidate-gene studies are often very satisfying; in terms of the complex genetics of interspecies variation, the candidate-gene technique has rarely made much headway.

Transcriptomic and molecular evolution surveys

The above examples are all cases where an evolution study was launched with a gene already known to govern a trait. In this section, we cover a complementary hypothesis-generating strategy in which there may be no previously known link between genotype and phenotype; instead, genomic or transcriptomic profiling reveals a gene with divergent molecular patterns between species, and authors go after it as a candidate determinant of species differences in a trait of interest.

Transcriptomics

In the modern era it can be straightforward and even fairly cheap to subject several species in a clade of interest to transcriptional profiling, and inspect expression divergences for those that could be relevant to a macroscopic phenotype. In one example, Abzhanov *et. al.* focused on divergence in beak size and shape between

species of Darwin's finches[9]. They mined transcriptomic data for genes whose expression differences between the species correlated with beak characteristics. Focusing on *Bmp4*, which encodes a growth factor expressed during avian development, the authors set up an experimental design for validation in which they could manipulate *Bmp4* expression levels in chickens via a retroviral vector system and evaluate their effects on morphology. The results revealed robust changes in beak size and shape in the expected direction. On the basis of this compelling finding, it can be argued that *Bmp4* expression mediates divergence in the beaks of these finches. However, from this study we know nothing about DNA sequence variants underlying species-level differences in either *Bmp4* expression or beak morphology. This sheds light on the key disadvantage of the transcriptomic approach: in the best case, it highlights components of a mechanism for trait divergence that likely act downstream of the causal, heritable genetic basis.

Molecular evolution

For a somewhat more direct strategy to guess at variants that underlie a trait difference between species, many genomicists make use of test paradigms that evaluate molecular measurements for evidence of natural selection. The usual setup is to focus on a species exhibiting an adaptive trait of interest, as a contrast with its extant relatives as representatives of the plausible ancestral state. In the focal species, sites in the genome showing signatures of positive selection are compelling candidate determinants of its divergent trait.

For this purpose, the field of molecular evolution has a wealth of test paradigms that operate on DNA sequence data from panels of species[10]. One intriguing example of a success story from these strategies is a study by Dong *et. al.* [11] of relatives of the potato blight pathogen *Phytophthora infestans,* which all exhibit different host specificities. The authors began with a set of 82 virulence determinants*,* in which an excess of amino acid changes relative to silent variants had been detected in one or more lineages of the clade. From this list, the authors decided to focus on *EpiC1*, a gene already known to target host defense proteases. The authors used *in vitro* assays to validate that EPIC1 proteins from *P. infestans* and its sister species *P. mirabilis* had different host protease targets, and their data also pinpointed a single amino acid that could almost entirely reproduce the alternate protease specificity. This was a remarkably clean molecular finding and points up the strength of the molecular-evolution framework, although any *in vivo* effect of the variation in *EpiC1* has not yet been established.

In another example, Prabhakar *et. al.*, interested in the effect of regulatory elements in the development of human-specific traits, developed their own test statistic to find regions in the human genome that have been rapidly diverging from other terrestrial vertebrates[12]. Specifically, their test was designed to find loci that are well conserved up until the human lineage, but have since accumulated a surprising number of variants. One interpretation of a signature of this kind is that the respective locus has been undergoing positive selection for an altered or new function. The authors identified

a 546-base pair element, *HACNS1*, which, in a mouse embryo model, drove the expression of a reporter in a derived pattern compared to chimpanzee or rhesus macaque. Strikingly, swapping just 13 base pairs in the human sequence with the chimpanzee sequence, or vice versa, was enough to modulate the expression phenotype. The *in vivo* effect of *HACNS1* is still unknown, although the authors speculate that it could play a vital role in limb and digit development.

A third instructive example is a recent study by Roop *et. al.* of sugar metabolism differences between *Saccharomyces* species[13]. This story started with an observation of polygenic *cis*-regulatory variation in yeast species across unlinked genes in the galactose metabolic pathway. The latter was used as input into a molecular-evolution-style test that operates on measurements of allele-specific expression in hybrids rather than DNA sequence, and reveals a signature of a change in selective pressure between the species[14]. The authors used allele-swap experiments to show that *cis*-regulatory regions at seven galactose genes were necessary and sufficient for a glucose specialization phenotype that distinguishes *S. cerevisiae* from sister species in the clade.

These molecular-evolution case studies make clear that even with no hints from the previous literature, signatures of selection between species can be a rich source of hypotheses about candidate determinants of their trait differences. One caveat, however, is that molecular-evolution tests for the most part don't incorporate phenotype data per se. Thus, a scan for loci under selection along a lineage of interest could hit upon determinants of any phenotype relevant for fitness in its niche; the importance of any one hit gene in any one trait can only be sorted out by experimental validation. Another disadvantage is that as standard approaches test one locus at a time for signatures of selection, owing to limited power, they will likely not find the complete genetic basis of a divergent trait of interest. Usually, a relatively small number of loci emerge from the test, and in the best case one or two will be validated as partial contributors to an interspecies difference. Thus, as with the candidate-gene framework we detailed above, the complexity of interspecies genetics has not been well understood from this literature.

Statistical genetics: direct mapping of genotype to phenotype

All the approaches we have covered so far generate hypotheses—some more suggestive than others—about plausible determinants of trait variation between species. We now cover the family of forward-genetic methods referred to as statistical genetics, which directly map DNA sequence variants causal for phenotypic differences between wild, outbred individuals. We discuss whether and how different flavors of this approach make sense for interspecific trait variation.

Classical linkage mapping

In the simplest statistical-genetic setup for the study of long-diverged species, a researcher would seek to map the variants that confer a derived, adaptive phenotype in

one species, relative to another species representing the putative ancestral state. This parallels the statistical-genetic framework known as linkage mapping, in which two genetically distinct parent individuals of the same species are crossed to generate recombinant progeny, among which variants across the genome are tested for co-inheritance with a trait of interest. The main hurdle blocking researchers from using this technique to study interspecies phenotypic variation is that, even if two species can be mated to produce viable offspring, it is very rare that this hybrid is fertile. The literature to date has made the most of these rare cases.

As an example of such a system, several groups have pursued the genetics of dark posterior abdomen pigmentation in male Drosophilids[15]. *Drosophila santomea* has lost the trait, and *D. yakuba* has retained the putatively ancestral pigmentation pattern. Carbone *et al.* used linkage mapping between these species to narrow down which genetic regions were important for the pigmentation trait difference and found strong signal for linkage to a locus on the X chromosome[16]. In a subsequent paper, Jeong *et. al.* focused on *tan*, a known pigmentation gene, as the best candidate in the linkage region[15]. In their validation experiments, a *D. melanogaster tan* transgene expressed in the *D. santomea* background was capable of partially restoring male pigmentation. Impressively, the authors further narrowed down the causal variants to two *cis*-regulatory nucleotides at the *tan* locus, and showed that the transformation of them into *D. santomea* was enough to rescue *tan* expression.

In another linkage mapping success story, Fridman *et. al.* used 76 introgression lines made with wild tomato (*Solanum pennellii*) alleles on a domesticated tomato background (*S. lycpersicum*) to map the genetic basis of a difference in sugar yield between the two species to a single region, *Brix9-2-5*[17]*.* Fine-mapping this locus pointed to *LIN5*, a cell-wall invertase as the likely cause of the pheotypic difference. To determine which variants within *LIN5* were responsible for the sugar yield phenotype, as well as an additional difference in *LIN5* enzyme activity between the species, the authors conducted another round of mapping using a variety of other *Solanum* species. They were able to narrow down the causal variant responsible for differences in *LIN5* sugar affinity to a single amino acid change in the coding region of *LIN5*.

These case studies demonstrate the utility of the linkage mapping workflow. The key upsides to this approach are that it directly maps loci underlying a trait of interest, and has the ability to pick up multiple different regions contributing to the trait, ultimately leading to insight into their differential contributions. The primary disadvantage of the linkage mapping paradigm for interspecific traits is simply that it's impossible for most species pairs, owing to reproductive isolation. Additionally, even in cases where interspecific hybrids are partially fertile, their inability to generate large broods can compromise the power of a linkage test owing to small sample size. Also, in many cases the broods that can be recovered will be biased in terms of which genotypes can be sampled, since certain combinations of alleles are often inviable.

New linkage mapping paradigms: recombination induced in vitro

Very recently, two powerful and creative approaches have been developed to circumvent the need for meiotic progeny, which renders linkage analysis so difficult in most pairs of species. Fundamentally, linkage analysis rests on a test for a relationship between genotype and phenotype among a panel of recombinant individuals from a cross between two parents; the idea in these new approaches is to artificially generate such recombinants outside the process of meiosis.

One of these two, Sadhu *et. al.,* leveraged the cell's natural ability to repair DNA damage during mitosis by homologous recombination[18]. It is possible to artificially induce mitotic recombination by targeting DNA damage, and as the authors realized, CRISPR-Cas9 is the perfect tool to accomplish this. Their proof of concept used variation between two strains of *S. cerevisiae*; in the hybrid formed from a mating between the strains, the authors introduced genomic cuts by Cas9 at each of 95 unique sites on yeast chromosome arm 7L, and used the resulting recombinant panel to map the genetic basis of variation in manganese resistance. The results revealed one amino acid polymorphism in an ion pump, *PMR1,* which contributed to but was not the sole determinant of the resistance trait. We include this case study in the current review even though this first application of this technique was on a trait difference within species. If anything, the approach should work even better in interspecific hybrids, which will harbor more genetic variation at which Cas9 can be targeted allele-specifically. One drawback to this technique is that every additional chromosome that would be targeted for Cas9-mediated recombination requires an increase in cost and complexity. Nevertheless, it is a scheme with great potential for interspecies genetics.

Another recently pioneered technique gets around the need for meiotic recombination in the generation of a panel of F2 individuals in a slightly different way. In this case, Lazzarano *et. al.* [19] took advantage of the drug ML216, which induces recombination during mitosis[20]. When treated with ML216, which inhibits the BLM DNA helicase, cells go through mitosis almost as normal with one exception: after DNA replication, recombination takes place at an elevated rate between the homologous chromosomes, thereby causing the exchange of alleles between the two parent species' genomes. After cytokinesis, two recombinant individuals are thus generated in the course of a single cell division. Lazzarano *et. al.* applied this technique to generate large panels of recombinants in interspecific hybrid embryonic stem cell lines. In an analysis of the divergent mouse species *Mus musculus* and *M. spretus*, the authors were able to map the genetic basis of sensitivity to the antimetabolite tioguanine (6-TG) to a region near to the gene hypoxanthine–guanine phosphoribosyltransferase (*Hprt*), an enzyme known to metabolize 6-TG. This scheme has the potential to map trait variation between species at nucleotide-level resolution and is remarkably fast relative to other alternatives in mice. The authors claim a single experiment could be conducted in as little as 6 days, whereas an equivalent experiment done in whole animals, even if capable of backcrossing, could take upwards of 1-2 years.

Reciprocal hemizygosity mapping

The most recent statistical-genetic approach relevant for interspecies variation relies on a scheme called the reciprocal hemizygosity test. In the classic version of this test, for a given gene, the phenotypic effect of each allele in turn in an interspecific hybrid diploid is uncovered via insertion of a transposon into the other copy, and the two transposon mutants are compared directly to reveal the impact of allelic variation on the trait. Our group developed a genome-scale pipeline for the reciprocal hemizygosity test which they called RH-seq, and used it to map eight loci governing differences in thermotolerance between yeast species[21]. Transgenesis experiments validated the effect of each as being sufficient or necessary, or both, for the thermotolerance phenotype, and molecular evolution tests revealed that the eight genes are undergoing accelerated evolution compared to the rest of the yeast genome. Interestingly, we found the phenotypic effect at these loci to be strongly background-dependent, as the effect sizes as measured in transgenesis experiments differed between strain backgrounds.

We consider RH-seq and ML216-mediated recombination as the closest the field has come to achieving unbiased genome-wide scans for the determinants of interspecies variation. Both have the limitation that they can only be used in species pairs that can be mated to form a viable hybrid. Apart from this constraint, each method has its own pros and cons. RH-seq can be applied in metazoans, whereas the *in vitro* recombination technique is dependent on ML216 administration to single cells. RH-seq tests the impact of allelic variation at the level of the gene; ML216-mediated approach could in theory have much finer resolution. We predict that each will accelerate progress in the field.

What have we learned?

Even the growing literature describing interspecific trait genetics, as we have reviewed it here, has barely scratched the surface in our understanding of how traits are built by evolution over long timescales. Nonetheless, some patterns have begun to emerge that can help guide the direction of future research.

• Subgenic determinants of species-level traits. Despite the focus on coding variants in the classic literature[22], these are not the sole mechanism for evolution between species. As in Roop *et. al.,* and Jeong *et. al.,* variation in the *cis*-regulatory regions of genes that modulate their expression can be a powerful player in complex traits. The emphasis on coding variants in early work in the field may be in part a product of acquisition bias, since many tests for molecular evolution at the sequence level focus on non-synonymous changes in coding regions.

• Complexity of species-level traits. Almost all of the studies detailed in this review paint the same picture: it is rare for only a single gene to be involved. In almost every case where transgenesis experiments have been performed on a candidate gene, it recapitulates only a fraction of the trait of interest, meaning other loci must be involved. Grappling with this prevalent genetic complexity will continue to complicate the dissection of both intra- and interspecies genetics. Despite the challenges, it is critical not to restrict focus in the field to genetically simple traits, their detectability and ease of

validation notwithstanding. Indeed, the architecture of ancient traits between species is particularly likely to be genetically complex, given that evolution has had such a long time for multiple mutations to accumulate.

• Epistasis in the architecture of species-level traits. It is becoming increasingly apparent that we cannot ignore the background dependence of allelic variation and epistasis between loci; again, this may become more and more of an issue the further apart in time two species have diverged.

In summary, each approach we have gone over in this review has its pros and cons. Allowing candidate genes to drive hypotheses has led to a large body of research on certain systems, but for the most part focuses on a single gene instead of the big picture. Yet focusing on a few genes has allowed some systems to be very deeply understood. QTL mapping studies in fertile interspecies hybrids have helped bridge the gap between genes and overall phenotypes, but small, limited, mapping populations have inhibited the ability of these studies to gain a complete understanding of the genetic architecture of the trait being studied. In our view, the future of interspecies genetics is in the unbiased, genome-wide dissection of traits. The best way to gain a full picture of how a trait has evolved is to systematically interrogate allelic differences between two species and ask, at each locus, whether natural variation influences the trait of interest or not. Endeavors of this nature have already begun to be published, with the advent of RH-seq, and ML216 and/or Cas9-mediated *in vitro* recombination. The main hurdles that studies of this kind will face are untangling the complexity of ancient traits and dealing with the effects of background dependence and epistasis. Although none of these problems are easy to solve, as more and more studies are completed, it will become easier to glean more basic principles about how new traits, and ultimately, species evolve.

References

1. Olson-Manning, C.F., Wagner, M.R. & Mitchell-Olds, T. Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat Rev Gen* **13(12)**, 867-877 (2012).
2. Savolainen, O., Lascoux, M. & Merilä, J. Ecological genomics of local adaptation. *Nat Rev Gen* **14**, 807-820 (2013).
3. Monteiro, A. & Podlaha, O. Wings, Horns, and Butterfly Eyespots: How Do Complex Traits Evolve? *PLoS Bio* doi:10.1371/journal.pbio.1000037 (2009).
4. Hoekstra, H.E. & Coyne, J.A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61(5)**, 995-1016 (2007).
5. Blount, Z.D., Borland, C.Z. & Lenski, R.E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli. PNAS* **105**, 7899-7906 (2008).
6. Daugherty, M.D., Schaller, A.M., Geballe, A.P. & Malik, H.S. Evolution-guided functional analyses reveal diverse antiviral specificities encoded by IFIT1 genes in mammals. *eLIFE* doi:10.7554/eLife.14228 (2016).
7. Campesan, S., Dubrova, Y., Hall, J.C. & Kyriacou, C.P. The nonA Gene in *Drosophila* Conveys Species-Specific Behavioral Characteristics. *Genetics* **158**, 1535-1543 (2001).
8. Palgrave, C.J. et. al. Species-Specific Variation in RELA Underlies Differences in NF-κB Activity: a Potential Role in African Swine Fever Pathogenesis. *Journal of Virology* **85(12)**, 6008-6014 (2011).
9. Abzhanov A., et. al. *Bmp4* and Morphological Variation of Beaks in Darwin's Finches. *Science* **305(5689)**, 1462-1465 (2004).
10. Vitti, J.J., Grossman, S.R. & Sabeti, P.C. Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet.* **47**, 97-120 (2013).
11. Dong, S. et. al. Effector Specialization in a Lineage of the Irish Potato Famine Pathogen. *Science* **343(6170)**, 552-555 (2014).
12. Prabhakar, S. et. al. Human-Specific Gain of Function in a Developmental Enhancer. *Science* **321(5894)**, 1346-1350 (2008).
13. Roop, J.I., Chang, K.C. & Brem, R.B. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature* **530(7590)**, 336-339 (2016).
14. Orr, H.A. Testing Natural Selection vs. Genetic Drift in Phenotypic Evolution Using Quantitative Trait Locus Data. *Genetics* **149**, 2099-2104 (1998).
15. Jeong, S. et. al. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* **132(5)**, 783-793 (2008).
16. Carbone, M.A., et. al. Quantitative Trait Loci Affecting the Difference in Pigmentation Between *Drosophila yakuba* and *D. santomea. Genetics* **171(1)**, 211-225 (2005).
17. Fridman, E. et. al. Zooming In on a Quantitative Trait for Tomato Yield Using Interspecific Introgressions. *Science* **305(5691)**, 1786-1789 (2004).
18. Sadhu, M.J., Bloom, J.S., Day, L. & Kruglyak, L. CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science* doi:10.1126/science.aaf5124 (2016).
19. Lazzarano, S. et. al. Genetic mapping of species differences via *in vitro* crosses in mouse embryonic stem cells. *PNAS* doi:10.1073/pnas.1717474115 (2018).

20. Nguyen, G.H. et. al. A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chem Bio* **20(1)**, 55-62 (2013).
21. Weiss, C.V. et. al. Genetic dissection of interspecific differences in yeast thermotolerance. *Nature Genetics* doi:10.1038/s41588-018-0243-4 (2018).
22. Stern, D.L. & Orgogozo, V. The loci of evolution: how predictable is genetic evolution? *Evolution* **62(9)**, 2155-2177 (2008).

Chapter 3

The genetic dissection of an ancient divergence in yeast thermotolerance

The contents of this chapter are based on the following publication, with permission from the authors:

Weiss, C.V., Roop, J.I., Hackley, R.K., Chuong J.N., Grigoriev, I.V., Arkin, A.P., Skerker, J.M., Brem, R.B. Genetic dissection of interspecific differences in yeast thermotolerance. *Nature Genetics*, in press (2018).

Abstract

Some of the most unique and compelling survival strategies in the natural world are fixed in isolated species[1]. To date, molecular insight into these ancient adaptations has been limited, as classic experimental genetics has focused on interfertile individuals in populations[2]. Here we use a new mapping approach, which screens mutants in a sterile interspecific hybrid, to identify eight housekeeping genes that underlie the growth advantage of *Saccharomyces cerevisiae* over its distant relative *S. paradoxus* at high temperature. Pro-thermotolerance alleles at these mapped loci were required for the adaptive trait in *S. cerevisiae* and sufficient for its partial reconstruction in *S. paradoxus*. The emerging picture is one in which *S. cerevisiae* improved the heat resistance of multiple components of the fundamental growth machinery in response to selective pressure. Our study lays the groundwork for the mapping of genotype to phenotype in clades of sister species across Eukarya.

Introduction

Geneticists since Mendel have sought to understand how and why wild individuals differ. Studies toward this end routinely test for a relationship between genotype and phenotype via linkage or association[2]. These familiar approaches, though powerful in many contexts, have an important drawback—they can only be applied to interfertile members of the same species. This rules out any case in which an innovation in form or function evolved long ago and is now fixed in a reproductively isolated population.

As organisms undergo selection over long timescales, their traits may be refined by processes quite different from those that happen early in adaptation[3,4]. We know little about these mechanisms in the wild, expressly because when the resulting lineages become reproductively incompatible, classic statistical-genetic methods cannot be used to analyze them[1]. To date, the field has advanced largely on the strength of candidate-based studies that implicate a single variant gene in an interspecific trait[5,6], with the complete genetic architecture often remaining unknown. Against the backdrop of a few specialized introgression[7-10] and molecular-evolution[11] techniques available in the field, dissection of complex trait differences between species has remained a key challenge.

Here we develop a new genetic mapping strategy, based on the reciprocal hemizygosity test[12,13], and use it to identify the determinants of a difference in high-temperature growth between isolated *Saccharomyces* yeast species. We validate the

contributions of the mapped loci to the thermotolerance trait, and we investigate their evolutionary history.

Results

Species differences in thermotolerance

At high temperature, the yeast *Saccharomyces cerevisiae* grows qualitatively better than other members of its clade[14-16], including its closest relative, *S. paradoxus*, from which it diverged ~5 million years ago[17]. In culture at 39°C, *S. cerevisiae* doubled faster than *S. paradoxus* and accumulated more biomass over a timecourse, a compound trait that we call thermotolerance. The magnitude of differences in thermotolerance between species far exceeded that of strain variation within each species (Figure 1), whereas no such effect was detectable at 28°C (Supplementary Figure 1). The failure by *S. paradoxus* to grow to high density at 39°C was, at least in part, a product of reduced survival relative to that of *S. cerevisiae,* as cells of the former were largely unable to form colonies after heat treatment (Supplementary Figure 2). In microscopy experiments, *S. paradoxus* cells were almost uniformly visible as large-budded dyads after 24 hours at 39°C (Supplementary Figure 3), suggestive of defects late in the cell cycle as a proximal cause of death[18]. No such tendency was apparent in *S. cerevisiae* at high temperature, or in either species at 28°C (Supplementary Figure 3).

Massively parallel reciprocal hemizygosity testing by RH-seq

We set out to dissect the genetic basis of *S. cerevisiae* thermotolerance, using a genomic implementation of the reciprocal hemizygote test[12,13] (Figure 2a). For this purpose, we first mated *S. cerevisiae* strain DBVPG1373, a soil isolate from the Netherlands, with *S. paradoxus* strain Z1, an English oak tree isolate. The resulting sterile hybrid had a thermotolerance phenotype between those of its purebred parents (Supplementary Figure 4). In this hybrid background we generated hemizygote mutants using a plasmid-borne, selectable PiggyBac transposon system[19]. We cultured the pool of mutants in bulk for ~7 generations at 39°C and, separately, at 28°C. From cells in each culture we sequenced transposon insertion locations[20] as a readout of the genotypes and abundance of mutant hemizygote clones present in the selected sample. In these sequencing data, at each of 4888 genes we detected transposon mutant clones in both species' alleles in the hybrid (Supplementary Figure 5), with transposon insertions distributed in a largely unbiased manner across the genome (Supplementary Figure 6). For a given gene, we tabulated the abundances of mutants whose transposon insertion fell in the *S. cerevisiae* allele of the hybrid, after high-temperature selection relative to the 28°C control, and we compared them to the abundance distribution of mutants in the *S. paradoxus* allele (Figure 2a). Any difference in abundance between these reciprocal hemizygote cohorts can be ascribed to variants between the retained alleles at the respective locus; we refer to the comparison as reciprocal hemizygosity analysis via sequencing (RH-seq). Integrating this approach with a quality-control pipeline (Supplementary Figure 5), in a survey of 3416 high-

coverage genes we identified 8 top-scoring hits (false discovery rate 0.01; Figure 2b). At each such locus, disruption of the *S. cerevisiae* allele in the hybrid was associated with low clone abundance after selection at 39°C relative to 28°C (Figure 2b), reflecting a requirement for the *S. cerevisiae* allele for thermotolerance. All of the genes mapped by RH-seq were annotated as housekeeping factors: *ESP1, DYN1, MYO1, CEP3, APC1,* and *SCC2* function in chromosome segregation and cytokinesis, and *AFG2* and *TAF2* in transcription/translation.

To evaluate the role in thermotolerance of genes that emerged from RH-seq, we first sought to verify that growth differences between hemizygotes at a given locus were the consequence of allelic variation, and not an artifact of our genomic approach. Toward this end, at each RH-seq hit gene we engineered hemizygotes by targeted deletion of each species' allele in turn in the hybrid. In growth assays, the strain lacking the *S. cerevisiae* allele at each gene grew poorly at high temperature (Figure 2b), with little impact at 28°C (Supplementary Figure 7), as inferred from RH-seq. Likewise, at each locus, the *S. paradoxus* allele made no contribution to the phenotype of the hybrid, since deleting it had no effect (Figure 2B and Supplementary Figure 7). Locus effect sizes from this single-gene validation paradigm largely paralleled the estimates from RH-seq ($R^2 = 0.74$). We conclude that RH-seq hits represent bona fide determinants of thermotolerance in the hybrid.

Validation of RH-seq gene hits in purebreds

We expected that variation at our RH-seq hits, though mapped by virtue of their impact in the hybrid, could also explain thermotolerance differences between purebred species. As a test of this notion, for each mapped gene in turn, we replaced the two copies of the endogenous allele in each purebred diploid with the allele from the other species. Growth assays of these transgenics established the *S. cerevisiae* allele of each locus as necessary or sufficient for biomass accumulation at 39°C, or both: thermotolerance in the *S. cerevisiae* background was compromised by *S. paradoxus* alleles at 7 of the 8 genes and, in *S. paradoxus*, improved by *S. cerevisiae* alleles at 6 of 8 loci (Figure 3). Allele replacements had little impact on growth at 28°C (Supplementary Figure 8). These trends mirrored the direction of locus effects from hemizygotes in the hybrid, though the magnitudes were often different. Most salient were the small effect sizes in *S. paradoxus* relative to other backgrounds, indicative of strong epistasis in this poorly-performing species (Supplementary Figure 9). Thus, the loci mapped by RH-seq in an interspecies hybrid contribute causally to thermotolerance in purebreds, with effect sizes that depend on the context in which they are interrogated.

Avid growth at high temperature is a defining characteristic of *S. cerevisiae* as a species, relative to other Saccharomycetes (refs. 14-16 and Figure 1). In principle, the loci mapped by RH-seq could be unique to the genetic architecture of thermotolerance in our focal *S. cerevisiae* strain, DBVPG1373, or be part of a mechanism common to many *S. cerevisiae* isolates. In support of the latter model, transgenesis experiments showed that a diverse panel of *S. cerevisiae* isolates all harbored alleles conferring modest but significant growth benefits at high temperature, and alleles from multiple *S.*

paradoxus isolates were deleterious (Supplementary Figure 10a-b). We detected no such impact at 28°C (Supplementary Figure 10a-b). Similarly, we found elevated sequence divergence from *S. paradoxus* to be a shared feature of *S. cerevisiae* strains at the loci mapped by RH-seq (using the absolute divergence measure D_{xy}; Supplementary Figure 10c). These findings indicate that the *S. cerevisiae* population accumulated divergent, pro-thermotolerance alleles at appreciable density in the loci mapped by RH-seq, consistent with a role in the trait for these genes across the species. And in the yeast phylogeny, RH-seq hit genes were distinguished by accelerated evolution along the branch leading to *S. cerevisiae,* as expected if the ancestral program has been conserved among the other species in the clade (Supplementary Figure 10c).

Discussion

In this work, we have developed the RH-seq method for genome-wide mapping of natural trait variation, and we have used it to elucidate the genetics of thermotolerance in reproductively isolated yeasts. Growth at high temperature is likely a derived character in *S. cerevisiae*[14-16], and the mechanism by which evolution built the trait, after the split from *S. paradoxus*, has remained unknown. In pursuing the genetics of this putative ancient adaptation, we complement studies of younger, intra-specific variants that erode thermotolerance, in the few *S. cerevisiae* isolates that have lost the trait relatively recently[12,21]. We have sought to shed light on more ancient evolutionary events by considering *S. paradoxus* as a representative of the ancestral state, to which thermotolerant *S. cerevisiae* can be compared.

Biological function of RH-seq hits

Using this approach, we have mapped eight loci at which *S. cerevisiae* alleles are necessary and sufficient for thermotolerance. As our RH-seq scan did not attain complete genomic coverage, the hits we did find likely represent a lower bound on the complexity of the architecture of the trait. Six of the RH-seq hit genes are essential for growth in standard conditions[22], and all eight contribute to fundamental growth processes. *ESP1, DYN1, CEP3, APC1, MYO1,* and *SCC2* mediate mitotic spindle assembly, chromatid cohesion and separation, cytokinesis, and mitotic exit; *AFG2* regulates the release of maturation factors from the ribosome; and *TAF2* encodes a TFIID subunit. In each case, our growth experiments in the interspecific hybrid have shown that the *S. paradoxus* allele acts as a hypomorph at high temperature. Our work leaves unanswered exactly how heat-treated *S. paradoxus* dies in the absence of these functions, though the cells' large-budded morphology strongly suggests regulated arrest or stochastic failure late in the cell cycle. That said, given that some but not all RH-seq hit loci have roles in mitosis, it is likely only one of the choke points at which *S. paradoxus* alleles are a liability at high temperature. Assuming that these heat-sensitive alleles also littered the genome of the common ancestor with *S. cerevisiae*, thermotolerance would have evolved along the *S. cerevisiae* lineage by resolving each of them, boosting the heat resistance of many housekeeping processes. Such a mechanism would dovetail with the recent finding that, across species, the limiting

temperature for cell growth correlates with the melting temperatures of a subset of essential proteins[23].

Pros and cons of RH-seq

These insights into the evolution of a complex yeast trait serve as a proof of concept for RH-seq. To date, the reciprocal hemizygosity test has led to landmark discoveries in a candidate-gene framework, confirming the effects of variation at a given locus identified by other means[12,13]. Schemes to scale up the test have generated a genome's worth of hemizygotes from deletion-strain purebreds, which tend to harbor secondary mutations that come through screens as false positives[24,25]. As such, a key advantage of RH-seq is that we carry out mutagenesis in the hybrid, which ensures coverage of essential genes and obviates the use of mutation-prone null genotypes. Furthermore, any secondary mutations that do arise in a given hemizygote clone, *e.g.* during a long competition in the condition of interest, would not have a strong influence on RH-seq mapping, because deep mutagenesis generates many independent clones per gene that are analyzed together. One important caveat of RH-seq, as in single-gene reciprocal hemizygote tests, is the assumption that no epistasis unique to the hybrid will mask the effects of loci underlying a trait difference of interest between the parents. In our case study, the genetic architecture of thermotolerance in the hybrid did bear out as relevant for the purebreds, albeit with locus effect sizes that varied across the backgrounds. More dramatic discrepancies may be particularly likely when the hybrid has a heterotic (*i.e.* extreme) phenotype and is a poor model for the genetics of the parents[26]. The choice of a non-heterotic hybrid in which to pursue RH-seq would be analogous to classical linkage mapping in a cross whose progeny have, on average, phenotypes that are intermediate between those of the parents.

In fact, although we have focused here on ancient divergence, the RH-seq method would be just as applicable to individuals within a species, as a high-resolution alternative to linkage analysis. We thus anticipate that RH-seq will accelerate the mapping of genotype to phenotype in many systems, whether the parents of a cross are closely related or members of a species complex that have been locally adapting for millions of years.

Materials and Methods

Strains

Strains used in this study are listed in Supplementary Table 1. Homozygous diploid strains of *S. cerevisiae* and *S. paradoxus* used as parents of the interspecific hybrid, and as the backgrounds for allele-swap experiments, were homothallic DBVPG1373 and Z1, respectively. In the case of the hybrid parents, each strain was rendered homozgyous null for *URA3* via homologous recombination with a HYGMX cassette, then sporulated; a given mated spore from a dissected tetrad was grown up into a diploid that was homozygous null at *URA3* and tested for the presence of both genomes by PCR with species-specific primers.

piggyBac transposon machinery

For untargeted, genome-scale construction of reciprocal hemizygotes in the *S. cerevisiae* x *S. paradoxus* hybrid, we adapted methods for piggyBac transposon mutagenesis[19] to develop a system in which the transposon machinery was borne on a selectable and counter-selectable plasmid lacking a centromere. We constructed this plasmid (final identifier pJR487) in three steps. In step I we cloned the piggyBac transposase enzyme gene driven by the *S. cerevisiae TDH3* promoter (from plasmid p3E1.2, a gift from Malcolm Fraser, Notre Dame) into plasmid pJED104 (which contains *URA3*, an ARS, and the *CEN6* locus, and was a gift from John Dueber, UC Berkeley). For this cloning, the amplification used a forward and reverse primer containing a BamHI and XhoI site, respectively, that upon restriction digest yielded sticky ends for ligation to recipient BamHI and XhoI sites in digested pJED104. We used the resulting plasmid as input into step II, removal of the *CEN6* sequence: we first amplified the entire plasmid with primers that initiated outside of *CEN6* and were directed away from it, and contained reciprocally complementary NheI sites; sticky ends of the linear PCR product were then ligated together for re-circularization. We used the resulting plasmid as input into step III, the cloning in of a construct comprised of the KANMX cassette flanked by long terminal arms (328bp and 361bp) from the piggyBac transposon. We first amplified KANMX from pUG6[27] and each transposon arm from p3E1.2, using primers that contained overlapping sequence on the fragment ends that would ultimately be the interior of the construct, and XbaI sites on the fragment ends that would ultimately be the 5' and 3'-most ends of the construct. We stitched the three fragments together by overlap extension PCR, digested the resulting construct and the plasmid from step II with XbaI, and annealed sticky ends of the two to yield the final pJR487 plasmid.

Untargeted hemizygote construction via transposon mutagenesis

For mutagenesis, pJR487 was gigaprepped using a column kit (Zymo Research) to generate ~11 mg plasmid. To prepare for transformation, JR507 (the *S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1 hybrid) was streaked from a -80°C freezer stock onto a yeast peptone dextrose (YPD, 1% yeast extract [BD], 2% yeast peptone [BD], 2% D-

glucose [Sigma]) agar plate and incubated for 2 days at 26°C. A single colony was inoculated into 100 mL YPD and shaken at 28°C, 200rpm for ~24 hours. The next day, we transferred cells from this pre-culture, and YPD, to each of four 1 L flasks at the volumes required to attain an optical density at 600 nm (OD_{600}) of 0.2 in 500 mL each. We cultured each for 6 hours at 28°C with shaking at 200rpm. Two of these cultures were combined into 1 L of culture and two into a separate 1 L, and each such culture was subjected to transformation (for a total of two transformations) as follows. The 1 L was split into twenty 50-mL conical tubes. Each aliquot was centrifuged and washed with water and then with 0.1 M lithium acetate (LiOAc, Sigma) mixed with 1X Tris-EDTA buffer (10 mM Tris-HCl and 1.0 mM EDTA); after spin-down, to each tube was added a solution of 0.269 mg of pJR487 mixed 5:1 by volume with salmon sperm DNA (Invitrogen), and then to each was added 3 mL of 39.52% polyethylene glycol, 0.12M LiOAc and 1.2X Tris-EDTA buffer (12 mM Tris-HCl and 1.2 mM EDTA). Tubes were rested for 10 minutes at room temperature, then heat-shocked in a water bath at 39°C for 26 minutes. Cells from all 20 tubes were then combined. We transferred cells from this post-transformation culture, and YPD, to each of three 1 L flasks at the volumes required to attain an OD_{600} of ~0.35-4 in 500 mL. Each such culture was recovered by shaking at 28°C and 200 rpm for 2 hours. G418 (Geneticin, Gibco) was added to each at a concentration of 300 µg/mL to select for those cells which had taken up the plasmid, and cultures were incubated with 200 rpm shaking at 28°C for two days until each reached an OD_{600} of ~2.3. All six such selected cultures across the two transformations were combined. We transferred cells from this combined culture, and YPD + G418 (300 ug/mL), to each of two 1 L flasks at the volumes required to attain an OD_{600} of 0.2 in 500 mL each. We cultured each flask at 28°C and 200 rpm shaking overnight until reaching an OD_{600} of 2.18 and combined the two cultures again to yield one culture. To cure transformants of the pJR487 URA+ plasmid, we spun down a volume of this master culture and resuspended in water with the volume required to attain a cell density of 1.85 OD_{600} units/mL. 12 mL of this resuspension were plated (1 mL per 24.1cm x 24.1cm plate) onto plates containing complete synthetic media with 5-fluoroorotic acid (5-FOA) [0.2% drop-out amino acid mix without uracil or yeast nitrogen base (YNB) (US Biological), 0.005% uracil (Sigma), 2% D-glucose (Sigma), 0.67% YNB without amino acids (Difco), 0.075% 5-FOA (Zymo Research)]. After incubation at 28°C to enable colony growth, colonies were scraped off all 12 plates and combined into water at the volume required to attain 40 OD_{600} units per 900 µL, yielding the final transposon mutant hemizygote pool. This was aliquoted into 1 mL volumes with 10% DMSO and frozen at -80°C.

Thermotolerance phenotyping via selection of the hemizygote pool

One aliquot of the pool of transposon mutant hemizygotes in the JR507 *S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1 hybrid background was thawed and inoculated into 150 mL of YPD in a 250 mL flask, and cultured for 7.25 hours at 28°C, with shaking at 200 rpm. We used this timepoint as time zero of our thermotolerance experiment, and took four aliquots of 6.43 mL (7 OD units) as technical replicates for sequencing of transposon insertion positions (see below). 9.19 mL of the remaining culture was back-diluted to an OD_{600} of 0.02 in a total of 500 mL YPD in each of six 2L

glass flasks for cultures that we call selections; three were grown at 28°C and three at 39°C (shaking at 200 rpm) until an OD_{600} of 1.9-2.12 was reached, corresponding to about 6.5 doublings in each case. Four cell pellets of 7 OD_{600} units each were harvested from each of these biological replicate flasks, for sequencing as technical replicates (see below). In total, 28 pellets were subjected to sequencing: 4 technical replicates from the time-zero culture; 3 biological replicates, 4 technical replicates each, from the 28°C selection; and 3 biological replicates, 4 technical replicates each, from the 39°C selection.

Tn-seq library construction

To determine the abundance of transposon mutant hemizygote clones after selection, we first sequenced transposon (Tn) insertions as follows. Each cell pellet from a time zero or selection sample (see above) was thawed on ice, and its genomic DNA (gDNA) was harvested with the ZR Fungal/Bacterial DNA MiniPrep kit (Zymo Research). gDNA was resuspended in DNA elution buffer (Zymo) pre-warmed to 65°C and its concentration was quantified using a Qubit 3.0 fluorometer. Illumina Tn-seq library construction was as described[28]. Briefly, gDNA was sonicated and ligated with common adapters, and for each fragment deriving from a Tn insertion in the genome, a sequence containing a portion of the transposon and a portion of its genomic context (the Tn-genome junction) was amplified using one primer homologous to a region in the transposon, and another primer homologous to a region in the adapter. The indexed adapter-specific primer was CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT, where the six N's are a unique index used for multiplexing multiple libraries onto the same Hiseq sequencing lane, and the transposon specific primer was ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNNNAGCAATATTTCAAGAATGCATGCGTCAAT, where N's are random nucleotides. Amplification used Jumpstart polymerase (Sigma) and the following cycling protocol: 94°C-2 min, [94°C-30 sec, 65°C-20 sec, 72°C-30 sec] X 25, 72°C-10 min. Sequencing of single-end reads of 150 bp was done over eight lanes on a HiSeq 2500 at the Joint Genome Institute (Walnut Creek, CA).

Tn-seq read-mapping and data analysis

For analysis of data from the sequencing of Tn insertion sites in pools of hemizygotes, we first searched each read for a string corresponding to the last 20 base pairs of the left arm of the piggyBac transposon sequence, allowing up to two mismatches. For each Tn-containing read, we then identified the genomic location of the sequence immediately downstream of the Tn insertion site, which we call the genomic context of the insertion, by mapping with BLAT (minimum sequence identity = 95, tile size = 12) against a hybrid reference genome made by concatenating amended *S. cerevisiae* DBVPG1373 and *S. paradoxus* Z1 genomes (see below). These genomic-context sequence fragments were of variable length; any case in which the sequence was shorter than 50 base pairs was eliminated from further analysis, as was any case in which a genomic-context sequence mapped to more than one location in the hybrid

reference. The resulting data set thus comprised reads containing genomic-context sequences specifically mapping to a single location in either *S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1, which we call usable reads. For a given library, given a cohort of usable reads whose genomic-context sequence mapped to the same genomic location, we inferred that these reads originated from clones of a single mutant with the Tn inserted at the respective site, which we call an insertion. In cases where the genomic-context sequences from reads in a given library mapped to positions within 3 bases of each other, we inferred that these all originated from the same mutant genotype and combined them, assigning to them the position corresponding to the single location to which the most reads mapped among those combined. For a given insertion thus defined, we considered the number of associated reads n_{insert} as a measure proportional to the abundance of the insertion clone in the cell pellet whose gDNA was sequenced. To enable comparison of these abundances across samples, we tabulated the total number of usable reads n_{pellet} from each cell pellet, took the average of this quantity across pellets, $\langle n_{pellet} \rangle$, and multiplied each n_{insert} by $\langle n_{pellet} \rangle / n_{pellet}$ to yield a_{insert}, the final normalized estimate of the abundance of the insertion clone in the respective pellet. For any insertions that were not detected in a given pellet's library ($n_{insert} = 0$) but detectable in another library of the data set, we assigned $n_{insert} = 1$.

We evaluated, from the mapped genomic-context sequence of each insertion, whether it fell into a gene according to the *S. cerevisiae* and *S. paradoxus* genome annotations[17,29], and we retained for further analysis only those insertions that fell into genes that were syntenic in the two species. For each such insertion, for each biological replicate corresponding to a selection culture (at 28°C or 39°C), we averaged the normalized abundances a_{insert} across technical replicates, yielding a single abundance estimate $\langle a_{insert} \rangle_{technical}$ for the biological replicate. We then calculated the mean of the latter quantities across all biological replicates of the selection, to yield a final abundance estimate for the insertion in this selection, $\langle a_{insert} \rangle_{total}$. Likewise, for each insertion and selection experiment we calculated $CV_{insert,total}$, the coefficient of variation of $\langle a_{insert} \rangle_{technical}$ values across biological replicates.

To use Tn-seq data in reciprocal hemizygosity tests, we considered for analysis only genes annotated with the same (orthologous) gene name in the *S. cerevisiae* and *S. paradoxus* reference genomes. For each insertion, we divided the $\langle a_{insert} \rangle_{total}$ value from the 39°C selection by the analogous quantity from the 28°C selection and took the $\log_2$ of this ratio, which we consider to reflect thermotolerance as measured by RH-seq. For each gene in turn, we used a two-tailed Mann-Whitney U test to compare thermotolerance measured by RH-seq between the set of insertions falling into the *S. cerevisiae* alleles of the gene, against the analogous quantity from the set of insertions falling into the *S. paradoxus* allele of the gene, and we corrected for multiple testing using the Benjamini-Hochberg method.

We tabulated the number of inserts and genes used as input into the reciprocal hemizygote test, and the number of top-scoring genes emerging from these tests, under each of a range of possible thresholds for coverage and measurement noise parameter values (Supplementary Figure 5). We used in the final analysis the parameter-value set

yielding the most extensive coverage and the most high-significance hits: this corresponded to insertions whose abundances had, in the data from at least one of the two selections (at 28°C or 39°C), $CV_{insert,total} \leq 1.5$ and $\langle a_{insert} \rangle_{total} \geq 1.1$, and genes for which this high-confidence insertion data set contained at least 5 insertions in each species' allele. This final data set comprised 110,678 high-quality insertions in 3416 genes. We used this complement of data in all display items of this paper with the following exception. To evaluate *post facto* the reproducibility across replicates of RH-seq measurements on genes called as hits, we first randomly paired each biological replicate at 39°C with one at 28°C; then, from the sequencing data from each pair in turn, we identified insertions whose abundances had $\langle a_{insert} \rangle_{total} \geq 1.1$ in at least one of the two temperatures for the respective replicate, and genes for which we had at least 5 such insertions in each species' allele. From these data, for a given RH-seq hit gene we tabulated single-replicate estimates of the abundances of hemizygotes harboring insertions in each species' homolog.

Amended reference genome construction

We generated reference genomes for *S. cerevisiae* strain DBVPG1373 and *S. paradoxus* strain Z1 as follows. Raw genome sequencing reads for each strain were downloaded from the SGRP2 database (see URLs). Reads were aligned using bowtie2[30] with default options; DBVPG1373 reads were aligned to version R64.2.1 of the reference sequence of the *S. cerevisiae* type strain S288C (Genbank Assembly Accession GCA_000146045.2), and Z1 reads were aligned to the *S. paradoxus* strain CBS432 reference sequence[31]. Single nucleotide variants (SNPs) were called using a pipeline of samtools[32], bcftools and bgzip, and were filtered for a quality score (QUAL) of >20 and a combined depth (DP) of >5 and either <65 (*S. cerevisiae*) or <255 (*S. paradoxus*). We then amended each reference genome with the respective filtered SNPs: we replaced the S288C allele with that of DBVPG1373 at each filtered SNP using bcftools' consensus command with default options (42,983 base pairs total), and amendment of the CBS432 sequence was carried out analogously using Z1 alleles (15,126 base pairs total).

Targeted-deletion hemizygote construction by homologous recombination

A given targeted hemizygote for each RH-seq hit gene except *TAF2* was generated in the *S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1 hybrid (JR507) by knocking out the allele of the gene from one species via homologous recombination with KANMX as described[37] with 70 base pairs of homology on the 5' and 3' ends of the cassette; checking was via diagnostic PCR. Two or more independent transformants were isolated and phenotyped for each hemizygote genotype (Supplementary Table 1).

Construction of allele replacement and targeted hemizgyote strains with Cas9

At each RH-seq hit gene, we constructed strains in wild-type homozygous diploid *S. cerevisiae* DBVPG1373 in which both copies of the endogenous allele were replaced by the allele from an *S. paradoxus* isolate (Z1 for Figure 3 and Supplementary Figures 8

and 9, and other strains as indicated for Supplementary Figure 10), and likewise for replacement of alleles from *S. cerevisiae* (DBVPG1373 in Figure 3 and Supplementary Figures 8 and 9, and other strains as indicated for Supplementary Figure 10) into *S. paradoxus* Z1. We call each such strain an allele-replacement strain, and each was constructed using a dual-guide Cas9 transgenesis method[38] in which a linear PCR fragment from the donor species is incorporated into the recipient genome by homology-directed repair of two chromosomal double-strand breaks induced by Cas9. Briefly, for each allele of each gene, we designed two guide RNAs for double-strand breaks by Cas9: one guide targeted a position ~1000 base pairs 5' to the coding start or at the 3' end of the closest upstream gene, whichever was closer, and the other guide targeted the region of the coding stop. The precise cut site of each was chosen to contain an NGG immediately downstream of variants between the *S. cerevisiae* and *S. paradoxus* strains, to avoid re-cutting of the donor allele by Cas9 after it had been introduced into the recipient strain. We cloned the two guide RNAs, a KANMX cassette, and the gene encoding the *S. pyogenes* Cas9 protein into a single plasmid as described[38]. The resulting plasmid was propagated in DH5a *E. coli* and miniprepped with a column miniprep kit (Qiagen). Separately, to generate the fragment to be used as the donor for DNA repair after Cas9 cutting, we PCR-amplified the respective region from the donor strain, with primers whose 5'-most 70 base pairs were homologous to the recipient and whose 3'-most 20 base pairs were homologous to the donor, except in the case of transgenesis using *S. cerevisiae* donors other than DBVPG1373 or *S. paradoxus* donors other than Z1 (for Supplementary Figure 10), for which the 3'-most 20 base pairs were homologous to DBVPG1373 or Z1, respectively. The homology region at the 5' end of the gene ended at most 31 base pairs upstream of the 5' cut site, and the homology region at the 3' end of the gene started at most 33 base pairs downstream of the 3' cut site. The donor fragment product was purified with a column kit (Qiagen) and ethanol-precipitated. We then simultaneously transformed, using the lithium acetate method, the donor fragment and dual-guide Cas9 plasmid into the recipient strain, using donor:acceptor ratios of 0.38:10 to 1:5, with 0.5-26 ug of plasmid. In this transformation, heat shock was for 20-30 minutes at 39-42°C in transformations of *S. cerevisiae* DBVPG1373, and 10-20 minutes at 37-39°C for transformations of *S. paradoxus* Z1. Transformants were plated on YPD+G418 (300 µg/mL) to select for cells that retained the plasmid. From this selection we patched single colonies onto YPD without G418, under the expectation that by the time a lawn came up for each patch, its cells would have lost the Cas9 plasmid. Each such strain was Sanger-sequenced at the junctions of the recipient and donor sequence. Positive patches were streaked to single colonies on YPD plates, and cells from each such colony were used to inoculate a patch on a YPD plate and, separately, to inoculate a patch on a YPD+G418 plate. Those colonies whose patches grew on the former but not on the latter were inferred to be cured of the plasmid and stored at -80°C. For all genes except *DYN1*, 2-3 such strains from each transformation were retained for thermotolerance assays and underwent Sanger sequencing of the entire locus to determine the exact swapped region (Supplementary Table 1). For *DYN1* allele-replacement in *S. paradoxus* Z1, the Cas9-based strategy yielded a single verified clone in which the *S. paradoxus* Z1 allele of *DYN1* was replaced by that of *S. cerevisiae* DBVPG1373, and likewise for *DYN1* allele-replacement in *S. cerevisiae* DBVPG1373. In each case, we mated the single swap

clone to a wild-type of the respective species background, confirmed heterozygosity of the resulting diploid via allele-specific diagnostic PCR at the *DYN1* locus, sporulated, and dissected tetrads, allowing each spore to autodiploidize and grow up as a homozygote; we retained from one such tetrad the two spores that were homozygous at the *DYN1* locus for the swapped allele, as confirmed by sequencing, and stored these at -80°C.

Targeted-deletion hemizygote strains for *TAF2* were generated by knocking out the *S. cerevisiae* or the *S. paradoxus* allele in the interspecific hybrid (JR507) using the above methods for Cas9 cutting and repair, with the following differences. To generate the fragment to be used as the donor for DNA repair after Cas9 cutting, we PCR-amplified the NATMX cassette from pBC713 (a gift from John Dueber, constructed as in [39]) using primers whose 5'-most 70 base pairs were homologous to the recipient and whose 3'-most 20 base pairs were homologous to the cassette. The homology region at the 5' end of the gene ended 22 base pairs outside the 5' cut site (which was upstream of coding start), and the homology region at the 3' end of the gene started 33 base pairs outside the 3' cut site. Positive strains were confirmed by PCR. Two independent transformants were isolated and phenotyped for each genotype (Supplementary Table 1).

Growth assays

Growth measurements of wild-type SGRP strains.

For the growth timecourse of a given wild-type, purebred homozygote isolate from the SGRP collection at 28°C, it was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. A single colony was inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm. This culture was back-diluted into YPD at an OD_{600} of ~0.05 and grown for an additional 5.5 hours at 28°C, 200 rpm, until reaching logarithmic phase. We transferred cells from each such pre-culture, and YPD, to each of 11 replicate wells of a 96-well plate, with volumes sufficient to yield a total volume of 150 µL per well at an OD_{600} of 0.02. The plate was covered with a gas-permeable membrane (Sigma) and incubated with orbital shaking in an M200 plate reader (Tecan, Inc.) at 28°C for 24 hours. For curves in Supplementary Figure 1a, measurements for optical density at 595nm (OD_{595}) were taken every 30 minutes and for each timepoint, the average was taken across replicate wells. To subtract background OD_{595} for the resulting curve, we tabulated the mean of the five lowest values from all datapoints, excluding the first two, and subtracted this value from that of each timepoint, setting any negative value to 0. To smooth the resulting curve, we first replaced each timepoint measurement by its average with those of the timepoints immediately before and after it; then, for any timepoint whose measurement was not greater than or equal to the previous one, we set it to be equal to that previous data point. For Supplementary Figure 1b, the efficiency for a given growth curve (from a single well) was calculated as the difference between the OD_{595} measured at the last four smoothed and averaged data points and that of the first four smoothed and averaged data points. Efficiencies from all of the wells from every *S. cerevisia*e isolate

were combined and compared to efficiencies from all of the wells for every *S. paradoxus* isolate in a two-sample two-tailed *t*-test.

For the growth timecourse of a given SGRP strain at 39°C (Figure 1A), we used a large-volume flask growth paradigm to avoid the influence of plate effects on growth measurements at high temperature in the incubated microplate reader, as follows. Each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. A single colony of a given strain was inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm. Each of these cultures was back-diluted into YPD at an OD_{600} of 0.05 and grown for an additional 5.5-7.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. We transferred cells from each such pre-culture, and YPD, to a glass 250 mL flask at the volumes required to attain an OD_{600} of 0.05 in 100 mL YPD, and incubated it at 39°C with shaking at 200 rpm. OD_{600} readings were taken every ~2 hours for ~18 hours. Figure 1A reports representative data from one of three such independent timecourse experiments. For curve fits, we used the getInitial and SSlogis functions in R to estimate starting values for the parameters of the logistic equation, and the nls function to fit the final parameters.

For efficiency measurements of a given SGRP isolate at 39°C in the large-volume format (Figure 1B), it was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. Two single colonies of each isolate were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate two replicate pre-cultures. Each was back-diluted into YPD at an OD_{600} at 600 nm of 0.05 and grown for an additional 5.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. The two pre-cultures were each again back-diluted into YPD in 1-inch diameter glass tubes with a target OD_{600} of 0.05; the actual OD_{600} of each was measured, after which it was grown at 39°C with shaking at 200rpm for 24 hours, and OD_{600} was measured again. The efficiency for each replicate was calculated as the difference between these final and initial OD_{600} values. The pipeline from inoculation off solid plates through pre-culture, two back-dilutions, and growth at 39°C we refer to as a day's growth experiment. For each day's experiments, we calculated the average efficiency across the replicates of each isolate $\langle e_{strain} \rangle$. We carried out two days' worth of replicate growth experiments for each isolate. For a given species, we used the complete cohort of measurements of $\langle e_{strain} \rangle$ from all isolates of each species, across all days, as input into a two-sample, two-tailed *t*-test to evaluate whether the suite of e_{strain} values across isolates of *S. cerevisiae* was significantly different from the analogous set of values from *S. paradoxus*.

Testing for heterosis in the DBVPG1373 x Z1 hybrid.

To compare efficiency at 39°C between the DBVPG1373 x Z1 hybrid and its *S. cerevisiae* and *S. paradoxus* parent strains (Supplementary Figure 4), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. Two single colonies of each isolate were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate two replicate pre-cultures. Each was back-diluted into YPD at an OD_{600} at 600 nm of 0.05 and grown for

an additional 5.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. The two pre-cultures were each again back-diluted into YPD in 1-inch diameter glass tubes with a target OD_{600} of 0.05; the actual OD_{600} of each was measured, after which it was grown at 39°C with shaking at 200rpm for 24 hours, and OD_{600} was measured again. The efficiency for each replicate, e_{strain}, was calculated as the difference between these final and initial OD_{600} values. The pipeline from inoculation off solid plates through pre-culture, two back-dilutions, and growth at 39°C we refer to as a day's growth experiment. We carried out two days' worth of replicate growth experiments for each isolate. To evaluate the significance of the difference between the growth of the hybrid and that of the *S. cerevisiae* parent, we used the complete cohort of measurements of e_{strain} from each strain, across all days, as input into a two-sample, two-tailed *t*-test.

Growth measurements of targeted-deletion hemizygotes and allele-replacement strains at 28°C.

For efficiency measurements of a given targeted-deletion hemizygote or allele-replacement strain at 28°C (Supplementary Figures 7, 8 and 10), pre-culture and plate reader assays were as for wild-type SGRP strains (see above), except that 6 or more replicate wells were cultured per strain. 2-3 independently isolated targeted-deletion hemizygotes or allele-replacement strains (Supplementary Table 1) were assayed for each genotype. Each timecourse of targeted-deletion hemizygote or allele-replacement strains also included the wild-type hybrid (JR507) or parent (*S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1), respectively, with pre-culture and replication as above. Efficiency for a given growth curve (from a single well) was calculated as the difference between the OD_{600} measured at the last four smoothed and averaged datapoints and that of the first four smoothed and averaged datapoints, with smoothing and averaging as detailed above. For Supplementary Figure 7, relative efficiency for a given well of a given targeted-deletion hemizygote strain at 28°C was tabulated as its efficiency divided by that of the average of all replicate wells of the wild-type hybrid (JR507) grown in the same experiment. For a given gene, we used the complete cohort of these measurements, from all isogenic hemizygotes, as input into a two-sample, two-tailed *t*-test to evaluate whether the relative efficiency of the strain in which the *S. cerevisiae* allele was knocked out was lower than the analogous quantity from the strain in which the *S. paradoxus* allele was knocked out. In Supplementary Figures 8 and 10, allele-replacement strains for a given gene were analyzed analogously, with the relative efficiency calculated against the respective wild-type parent, and with a one-sample, two-tailed *t*-test to evaluate whether the relative efficiency was significantly different from 1.

Growth measurements of targeted-deletion hemizygotes and allele-replacement strains at 39°C.

For efficiency measurements of a given targeted-deletion hemizygote strain at 39°C in the large-volume format (Figure 2B), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. Two single colonies of a given strain were each inoculated into liquid YPD and grown for 24 hours

at 28°C with shaking at 200 rpm. Each such pre-culture at stationary phase, or a log-phase outgrowth of it (which we used in the case of *DYN1* and *TAF2*: the pre-culture and YPD were added at the volumes required to attain an OD_{600} of 0.05 and grown for an additional 5.5 hours at 28°C, shaking at 200 rpm, until the culture reached logarithmic phase) was used to inoculate YPD in 1-inch diameter glass culture tubes with a target cell density corresponding to an OD_{600} of 0.05. The actual OD_{600} of each was measured, after which it was grown at 39°C with shaking at 200 rpm for 24 hours, and OD_{600} was measured again. The efficiency for each such replicate was then calculated as the difference between the final and initial OD_{600} values. The pipeline from inoculation off solid plates through pre-culture, back-dilution, and growth at 39°C we refer to as a day's growth experiment for a targeted-deletion homozygote. In each such experiment, 2-3 independently isolated targeted-deletion hemizygotes of a given gene in each direction were all assayed on the same day, alongside the wild-type hybrid parent (JR507) with replicate structure and methods as above. From each day's experiments, we calculated the average efficiency across the replicates of the wild-type hybrid $\langle e_{hybrid} \rangle$, and we used this quantity to normalize the efficiency $e_{hemizyg}$ measured for each replicate of each hemizygote strain assayed on that day. Thus, the final observable used for analysis for each replicate on a given day was $e_{hemizyg}/\langle e_{hybrid} \rangle$. We carried out 2-3 days' replicate growth experiments for each gene's hemizygotes. For a given gene, we used the complete cohort of these measurements of $e_{hemizyg}/\langle e_{hybrid} \rangle$, from all days and all isogenic hemizygotes, as input into a two-sample, one-tailed *t*-test to evaluate whether $e_{hemizyg}/\langle e_{hybrid} \rangle$ of the strain in which the *S. cerevisiae* allele was knocked out was lower than the analogous quantity from the strain in which the *S. paradoxus* allele was knocked out. To evaluate the agreement between the effect size of variation at a given locus in these experiments and the estimate from RH-seq data, we first processed the latter as follows. For a given insert in the RH-seq data set we defined thermotolerance as the ratio between $\langle a_{insert} \rangle_{total}$ after growth at 39°C and $\langle a_{insert} \rangle_{total}$ after growth at 28°C. We calculated the average thermotolerance across all inserts passing the cutoffs delineated in Supplementary Figure 5 that reflected hybrid strains harboring a wild-type copy of the allele from *S. paradoxus* Z1 of the focal gene and a transposon insertion in the *S. cerevisiae* DBVPG1373 allele; separately, we calculated the analogous quantity from hybrids with an insertion in the *S. paradoxus* allele. The ratio of these two thermotolerance values we refer to as the effect size for the locus of interest as measured in RH-seq. Next, to quantify the effect size for a given locus from our deletion-based hemizygosity experiment, we tabulated the average $e_{hemizyg}/\langle e_{hybrid} \rangle$, for all isogenic hybrid strains harboring a wild-type copy of the allele from *S. paradoxus* Z1 of the focal gene and a deletion of the *S. cerevisiae* DBVPG1373 allele; separately, we calculated the analogous quantity from strains in which the *S. paradoxus* allele was knocked out. The ratio between these two averages we consider the effect size from deletion-based hemizygotes. We used these two cohorts of effect size values (from RH-seq and deletion-based hemizygosity experiments) as input into a Pearson correlation calculation across all eight loci.

For growth measurements of a given allele-replacement strain at 39°C in the large-volume format (Figure 3, Supplementary Figures 7, 8 and 10), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3

days. 1-2 single colonies of each strain were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate 1-2 replicate pre-cultures. Each was back-diluted into YPD at an OD_{600} of 0.05 and grown for an additional 5.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. Each pre-culture were each again back-diluted into YPD in 1-inch diameter glass tubes with a target OD_{600} of 0.05 (for experiments using a single pre-culture, it was now split into two replicate pre-cultures, each of the same OD_{600}); the actual OD_{600} of each was measured, after which it was grown at 39°C with shaking at 200rpm for 24 hours, and OD_{600} was measured again. The efficiency for each replicate was calculated as the difference between these final and initial OD_{600} values. The pipeline from inoculation off solid plates through pre-culture, two back-dilutions, and growth at 39°C we refer to as a day's growth experiment for an allele-swap strain. In each such experiment, 2-3 independently isolated allele-swap strains targeting a given gene in a given background were all assayed on the same day, alongside the respective wild-type background strain (*S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1) with replicate structure and methods as above. For each day's experiments, we calculated the average efficiency across the replicates of the wild-type parent $\langle e_{parent} \rangle$, and we used this quantity to normalize the efficiency e_{swap} measured for each replicate assayed on that day of each allele-swap strain in the respective background. Thus, the final measurement used for analysis for each replicate on a given day was $e_{swap}/\langle e_{parent} \rangle$. We carried out 2-3 days' worth of replicate growth experiments for each gene's allele-swap strains. For a given gene in a given background, we used the complete cohort of measurements of $e_{swap}/\langle e_{parent} \rangle$ from all days and all allele-swap strains as input into a one-sample, one-tailed *t*-test to evaluate whether $e_{swap}/\langle e_{parent} \rangle$ was significantly different from 1. For swaps of the *S. cerevisiae* allele of a given gene into *S. paradoxus*, we tested whether $e_{swap}/\langle e_{parent} \rangle$ was greater than 1 (*i.e.* that the swap strain grew better at 39°C than did its parent), and for swaps of the *S. paradoxus* allele of a given gene into *S. cerevisiae*, we tested whether $e_{swap}/\langle e_{parent} \rangle$ was less than 1 (*i.e.* that the swap strain grew worse at 39°C than its parent).

Testing survival of wild-type parent strains at 39°C.

To test the survival of heat-treated *S. paradoxus* Z1 and *S. cerevisiae* DBVPG1373 (Supplementary Figure 2), each strain was streaked from -80°C freezer stocks onto YPD agar plates and incubated at 26°C for 3 days. 1-2 single colonies of each parent were inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to create 1-2 replicate pre-cultures. After 24 hours, we transferred cells from each pre-culture, and YPD, to each of 2-4 tubes at the volumes required to attain an OD_{600} of 0.05 in 11 mL YPD. Two tubes were incubated at 28°C and two at 39°C, all with shaking at 200 rpm for 24 hours. OD_{600} of each was measured; for each culture grown at 28°C, 100 µL of a 1.0×10^{-5} serial dilution was plated to YPD, and for cultures grown at 39°C, *S. paradoxus* and *S. cerevisiae* were serially diluted to 10^{-1} and 5.0×10^{-5}, respectively. Plates were incubated at 26°C for three days until single colonies appeared. Colonies on each plate were counted, from which we tabulated the colony forming units per mL of culture plated, and normalized by the optical density of the original plated culture. The pipeline from pre-culture through treatment and colony

counting we refer to as one day's worth of experiments. We used the results of two days' worth of experiments (a total of four for each species and temperature) as input into an ANOVA with species and temperature as factors, and took the *p*-value for the interaction between the factors as the estimate of the significance of the difference between species in the effect of temperature (of the 24-hour liquid culture) on cells' ability to grow into colonies (at the permissive temperature).

Microscopy and quantification

For microscopy of *S. cerevisiae* DBVPG1373 and *S. paradoxus* Z1 (Supplementary Figure 3), each species was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. Two single colonies of each strain were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate two replicate pre-cultures. Each was then back-diluted into YPD at an OD_{600} of 0.05; one was grown at 39°C with shaking at 200rpm for 24 hours, and the other was grown at 28°C with shaking at 200rpm for 24 hours. After the 24 hour growth period, 0.5 OD units of each culture were harvested through centrifugation and incubated in 66.5% ethanol for 1-4 hours at room temperature. Each sample was washed twice with 1X Dulbecco's Phosphate Buffered Saline (DPBS, Gibco), resuspended in 0.5 mL of 1X DPBS, and vortexed for 15 seconds on high. 5 µL of each sample was transferred to an agarose pad made with 1% agarose and YPD. We observed samples on a Zeiss Axio Observer inverted bright-field microscope at 100X magnification. Images were taken using a Hamamatsu ORCA-Flash4.0 digital camera and visualized using ZEN software for image analysis. The exposure of each image was set automatically through ZEN, and brightness was adjusted using the "Min/Max" adjustment for black and white light. The pipeline from inoculation off solid plates through pre-culture, growth at 39°C or 28°C, fixation, and imaging we refer to as a day's experiment. We carried out 2 days' worth of experiments for each species, yielding a total of 17-21 images at each temperature, for each species. In each image, free-floating cells were scored manually as singlets or those with a small, medium, or large bud (for a total of 31-151 scored cells per species and temperature). The proportion of scored cells with large buds was tabulated for each day's experiment for each species and temperature.

Locus effect sizes

Locus effect sizes in Supplementary Figure 9 were computed from the data in Figures 2B and 3 of the main text as follows. For analyses in the hybrid background, for a given locus we calculated $m_{S.par}$, the mean of all replicate measurements of $e_{hemizyg}/\langle e_{hybrid}\rangle$ of hemizygote strains lacking the *S. paradoxus* allele. We took, as independent estimates of the locus effect size (each corresponding to a circle on the respective grey bar in Supplementary Figure 9), each measurement of $e_{hemizyg}/\langle e_{hybrid}\rangle$ of a hemizygote strain lacking the *S. cerevisiae* allele, as a ratio against $m_{S.par}$. For analyses in the *S. cerevisiae* DBVPG1373 background, for a given locus we used each measurement of $e_{swap}/\langle e_{parent}\rangle$ as an independent estimate of the locus effect size (each corresponding to a circle on the respective orange bar in Supplementary Figure 9). For analyses in the *S. paradoxus* Z1 background, for a given locus we used each

measurement of $\langle e_{parent} \rangle / e_{swap}$ as an independent estimate of the locus effect size (each corresponding to a circle on the respective blue bar in Supplementary Figure 9).

Sequence analysis

D_{xy} *analysis*.

To evaluate whether sequence divergence from *S. paradoxus* at RH-seq hit loci was a shared feature of *S. cerevisiae* isolates, we used the D_{xy} statistic[33], the average number of pairwise differences between *S. cerevisiae* strains and *S. paradoxus*, normalized for gene length, as follows. We downloaded *S. cerevisiae* genomic sequences from the following sources: YJM978, UWOPS83-787, Y55, UWOPS05-217.3, 273614N, YS9, BC187, YPS128, DBVPG6765, YJM975, L1374, DBVPG1106, K11, SK1, 378604X, YJM981, UWOPS87-2421, DBVPG1373, NCYC3601, YPS606, Y12, UWOPS05-227.2, and YS2 from the Yeast Resource Center (see URLs); Sigma1278b, ZTW1, T7, and YJM789 from the Saccharomyces Genome Database (see URLs); and RM11 from NCBI (accession PRJNA13674). For each strain, we extracted the coding sequence of each gene in turn, and we downloaded the *S. paradoxus* reference sequence for each orthologous coding region from [17]. Sequences were aligned using MAFFT[34] with default settings. Alignments that did not contain a start and stop codon, or those that contained gaps at greater than 40% of sites were considered poor quality and discarded. We tabulated D_{xy} for each gene. To evaluate whether the eight RH-seq hit genes were enriched for elevated D_{xy}, we first tabulated $\langle D_{xy} \rangle_{true}$, the mean value across the eight RH-seq hit genes. We then sampled eight random genes from the set of 3416 genes tested by RH-seq; to account for biases associated with lower rates of divergence among essential genes, the resampled set contained six essential genes and two non-essential genes, mirroring the breakdown of essentiality among the RH-seq hits. Across this random sample we tabulated the mean D_{xy}, $\langle D_{xy} \rangle_{resample}$. We repeated the resampling 5000 times and used as an empirical *p*-value the proportion of resamples at which $\langle D_{xy} \rangle_{resample} \leq \langle D_{xy} \rangle_{true}$.

Phylogenetic analysis.

We downloaded orthologous protein coding regions for the type strains of *S. cerevisiae*, *S. paradoxus*, and an outgroup, *S. mikatae*, from [17]. For each gene for which ortholog sequences were available in all three species, we aligned the sequences with PRANK[35] utilizing the "-codon" option for codon alignment. These alignments were used as input into the codeml module of PAML[36], which was run assuming no molecular clock and allowing omega values to vary for each branch in the phylogeny. From the resulting inferences, we tabulated the branch length on the *S. cerevisiae* lineage for each gene. To evaluate whether sequence divergence of the eight RH-seq hit genes showed signatures of rapid evolution along the *S. cerevisiae* lineage, we used the resampling test detailed above.

Code availability

Custom Python and R scripts used for RH-seq data analysis are available on Github (see URLs).

URL list

SGRP2 Database - ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/SGRP2/input/strains
Yeast Resource Center - http://www.yeastrc.org/g2p/home.do
Saccharomyces Genome Database - http://www.yeastgenome.org/
RH-seq data analysis scripts - https://github.com/weiss19/rh-seq

Acknowledgements

We thank F. AlZaben, A. Flury, G. Geiselman, J. Hong, J. Kim, M. Maurer, and L. Oltrogge for technical assistance, D. Savage for his generosity with microscopy resources, and B. Blackman, S. Coradetti, A. Flamholz, V. Guacci, D. Koshland, C. Nelson, and A. Sasikumar for discussions; we also thank J. Dueber (Department of Bioengineering, UC Berkeley) for the PiggyBac plasmid. This work was supported by R01 GM120430-A1 and by Community Sequencing Project 1460 to RBB at the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility. The work conducted by the latter was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Figures

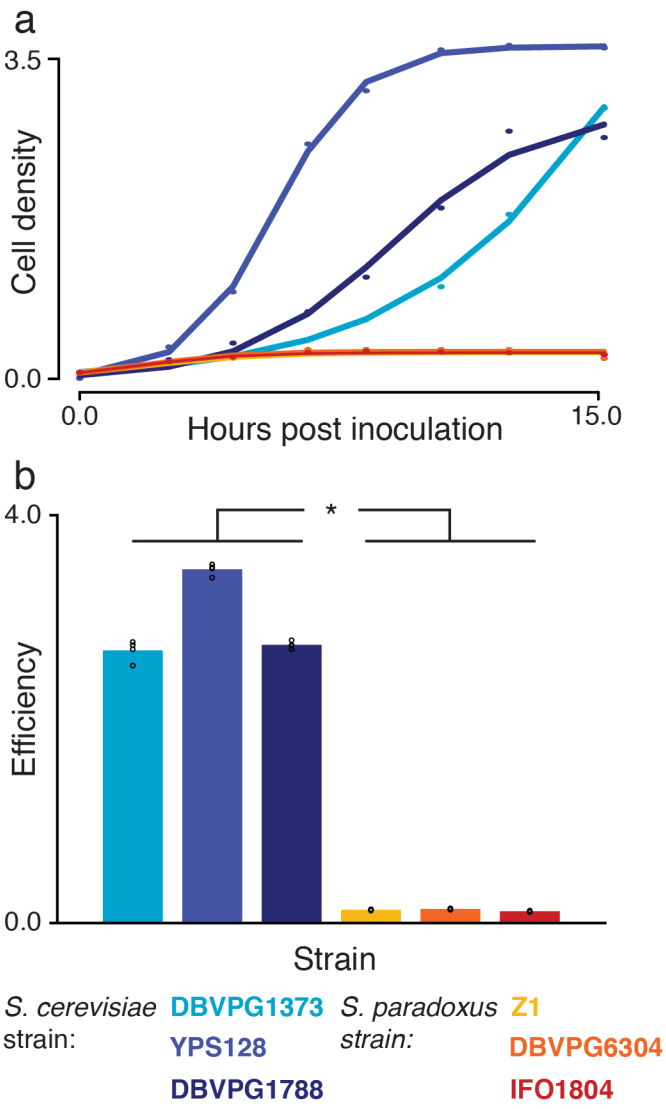


Figure 1. *S. cerevisiae* grows better at high temperature than *S. paradoxus*. **a**, Each point reports cell density (OD_{600}) of the indicated wild isolate of *S. cerevisiae* (blue) or *S. paradoxus* (orange) over a timecourse of growth at 39°C. Each solid line shows a logistic population growth model fit to the respective cell density measurements. **b**, Each bar reports mean efficiency (n = 4 cultures) of the indicated strain at 39°C, defined as the difference between cell density at 24 hours of growth and that at the time of inoculation. *, $p = 3.78 \times 10^{-11}$, two-sample, two-tailed t-test; individual measurements are reported as circles.

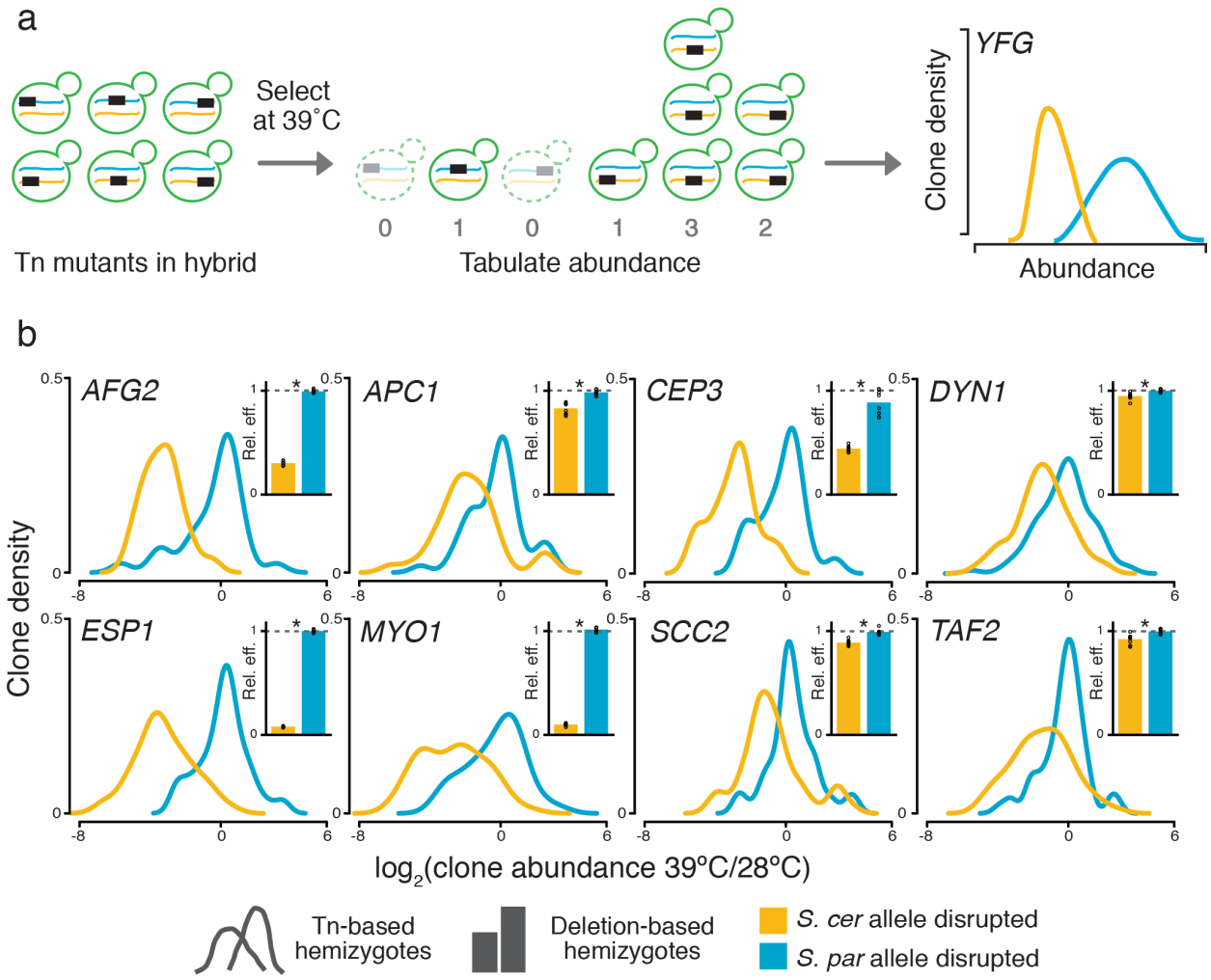


Figure 2. Mapping thermotolerance by RH-seq.
a, A transposon (Tn, rectangle) disrupts the allele from *S. cerevisiae* (blue) or *S. paradoxus* (orange) of a gene (*YFG*) in an interspecific hybrid (green). Clones lacking the pro-thermotolerance allele grow poorly at 39°C (dashed outlines), as measured by sequencing and reported in smoothed histograms (traces, colored to indicate the species' allele that is not disrupted). **b**, Each panel reports results from one RH-seq hit locus. In the main figure of a given panel, the *x*-axis reports the $\log_2$ of abundance, measured by RH-seq after selection at 39°C, of a clone harboring a transposon insertion in the indicated species' allele, relative to the analogous quantity for that clone from selection at 28°C. The *y*-axis reports the proportion of all clones bearing insertions in the indicated allele that exhibited the abundance ratio on the *x*, as a kernel density estimate. In insets, each bar reports the relative efficiency, calculated as the mean growth efficiency at 39°C (n = 8-12 cultures) of the indicated targeted-deletion hemizygote measured in liquid culture assays, normalized to the analogous quantity for the wild-type hybrid. *, $p \leq 0.002$, two-sample, one-tailed *t*-test; individual measurements are reported as circles.

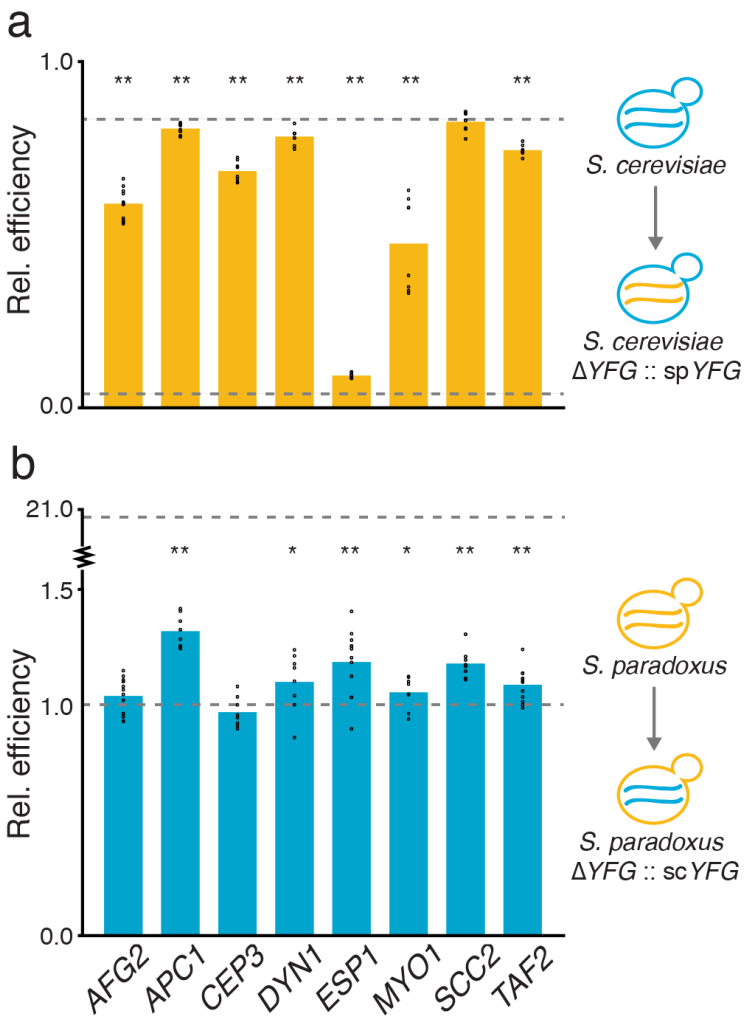
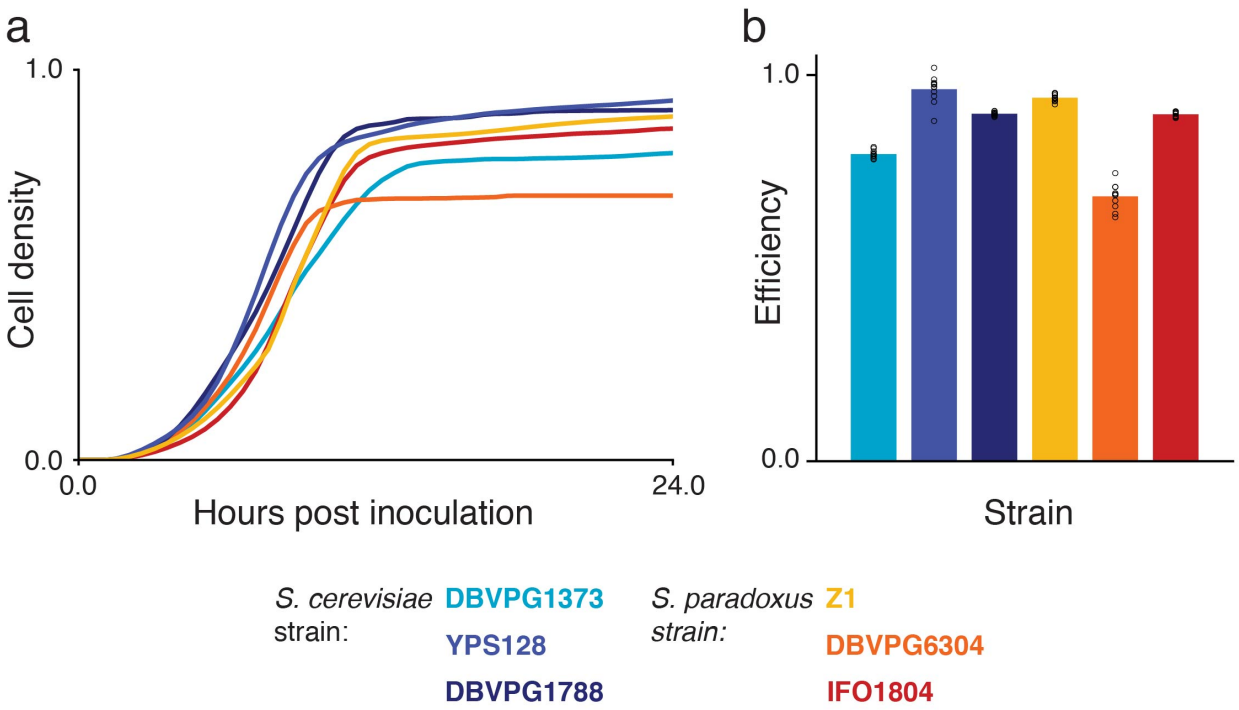
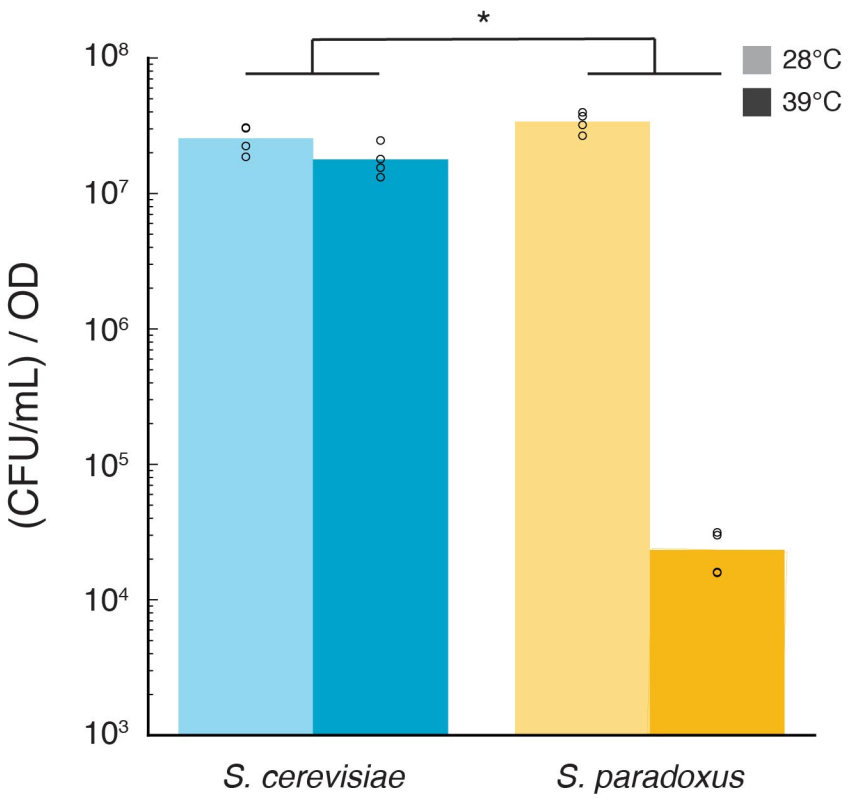


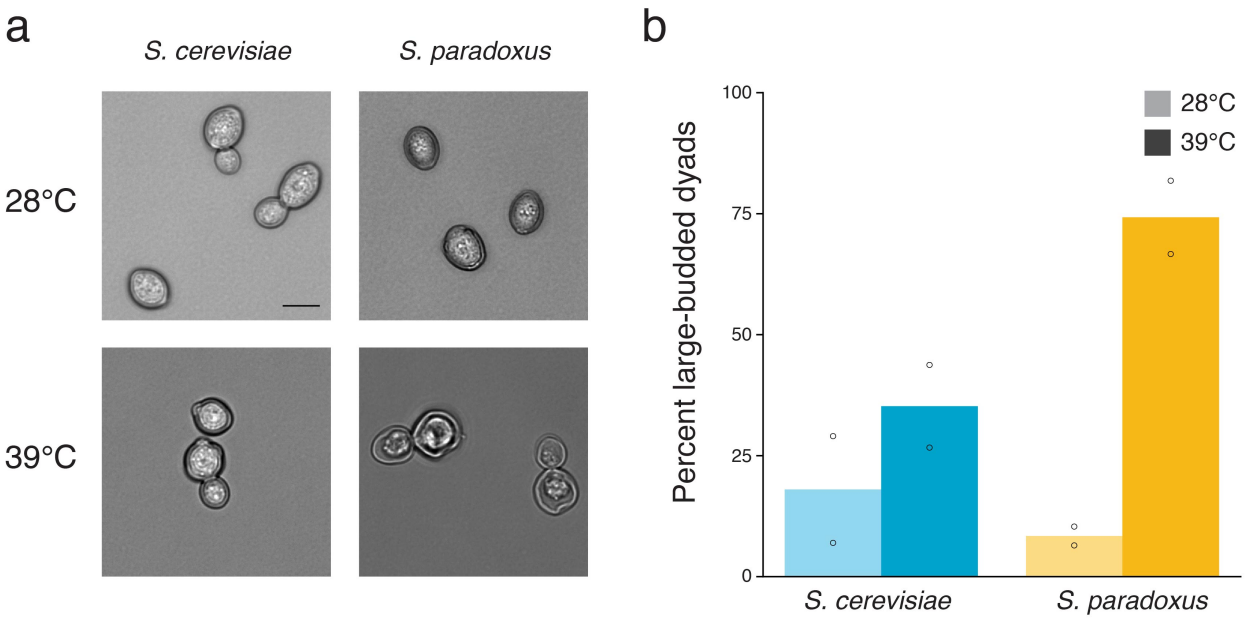
Figure 3. *S. cerevisiae* thermotolerance alleles are necessary and sufficient for growth at high temperature.
a, Each bar reports mean growth efficiency at 39°C, measured in liquid culture assays (*n* = 8-12 cultures), of an *S. cerevisiae* strain harboring the *S. paradoxus* allele at the indicated RH-seq hit locus, relative to the analogous quantity for wild-type *S. cerevisiae*. **b**, Data are as in **a**, except that each bar reports results from a *S. paradoxus* strain harboring the *S. cerevisiae* allele at the indicated locus, normalized to wild-type *S. paradoxus*. In a given panel, the top and bottom dotted lines report the relative efficiency of wild-type *S. cerevisiae* and *S. paradoxus*, respectively. *, $p \leq 0.036$; **, $p \leq 0.001$, one-sample, one-tailed *t*-test; individual measurements are reported as circles.



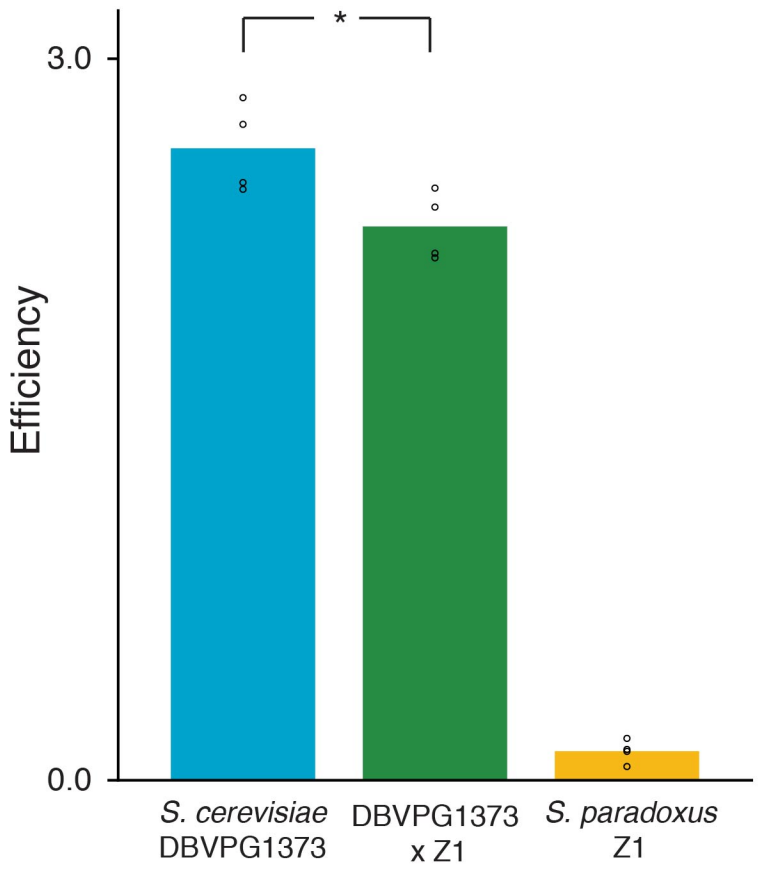
Supplementary Figure 1. *S. cerevisiae* and *S. paradoxus* do not differ significantly with respect to growth at 28°C.
a, Each trace reports mean optical density (OD_{595}) over time of the indicated wild isolate of *S. cerevisiae* (blue) or *S. paradoxus* (orange) cultured at 28°C (n = 11 cultures). **b**, Each bar reports the mean efficiency (n = 11 cultures) of the indicated strain after 24 hours of growth at 28°C; individual measurements are reported as circles. Efficiencies across strains were not significantly different between the species (p = 0.07, two-sample, two-tailed t-test).



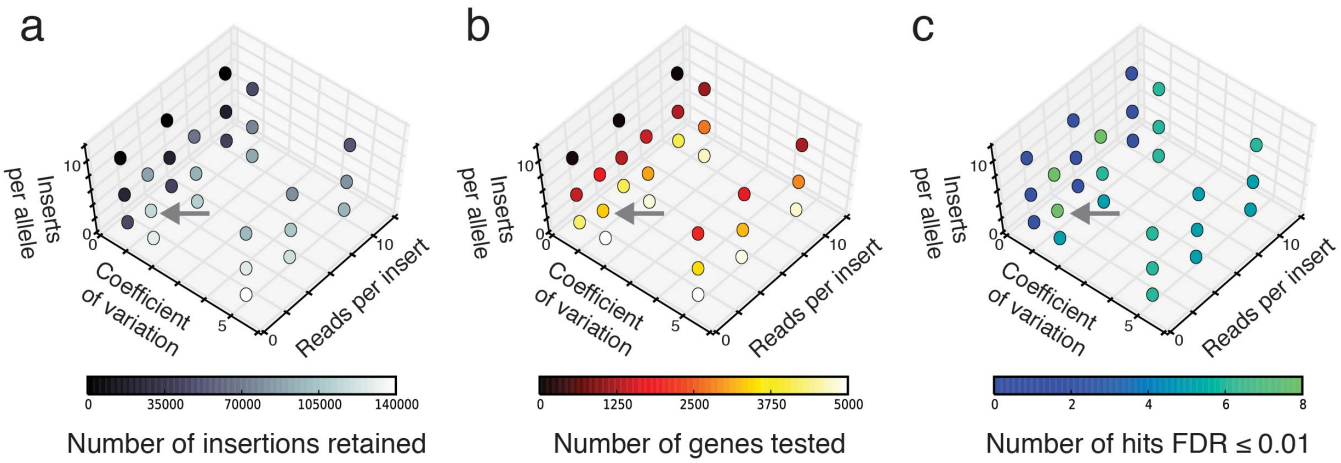
Supplementary Figure 2. Survival of *S. cerevisiae* cells is higher than that of *S. paradoxus* at 39°C.
Each pair of bars reports the mean number of colony-forming units, per mL of culture per unit of cell density (OD_{600}), from a liquid culture of *S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1 grown for 24 hours at the indicated temperature (*n* = 4 cultures). *, interaction *p* = 0.000152 in an ANOVA with species and temperature as factors; individual replicates are reported as circles.



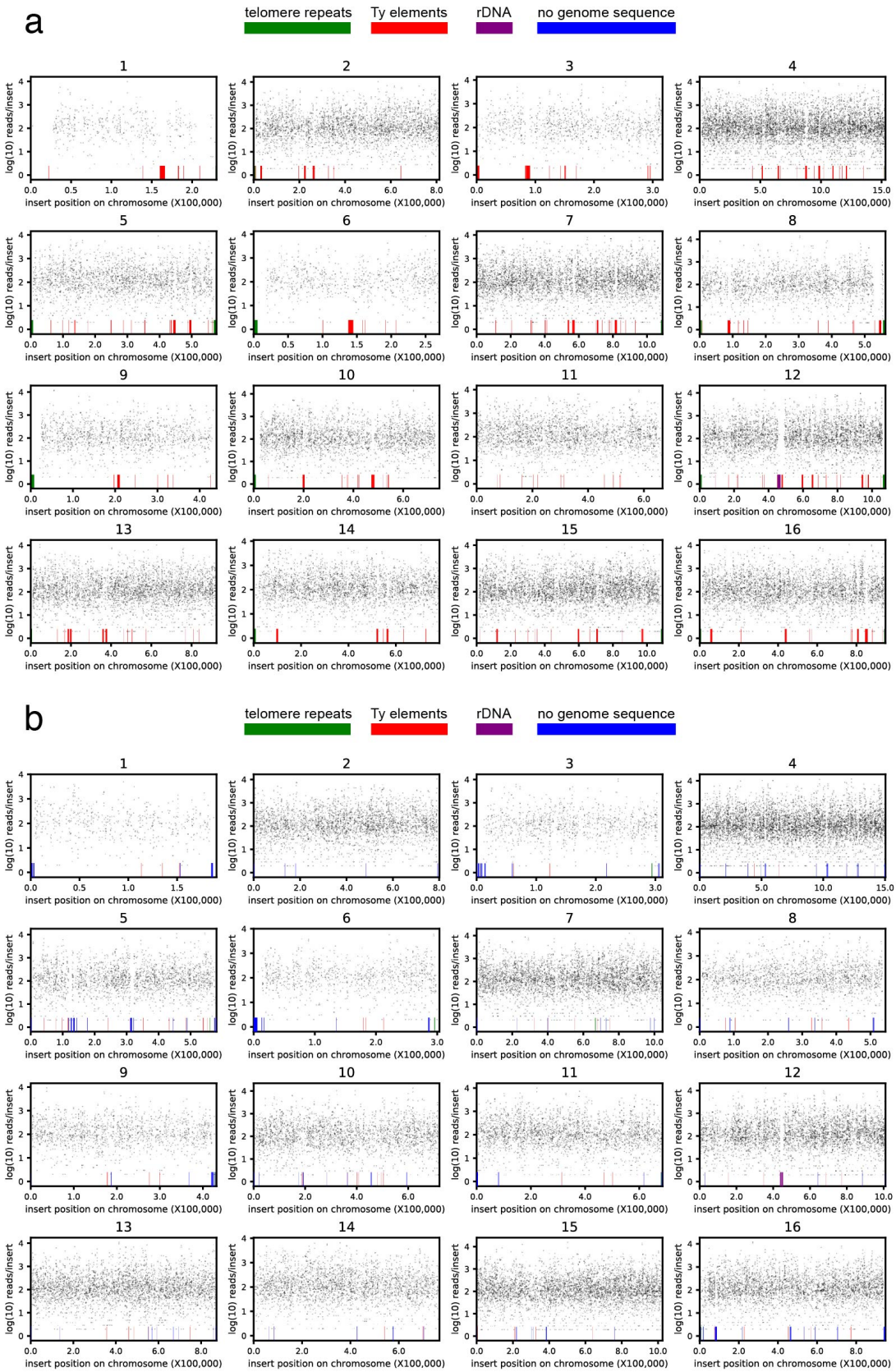
Supplementary Figure 3. *S. paradoxus* cells are predominantly large-budded dyads at high temperature.
Each panel reports results from microscopy experiments of *S. cerevisiae* DBVPG1373 and *S. paradoxus* Z1 after 24 hours of liquid growth at the indicated temperature; at 28°C, both species were approaching stationary phase. **a,** Representative images. Scale bar, 5 μm. **b**, Each pair of bars reports the mean proportion of large-budded dyads (*n* = 2 cultures), for the indicated species and temperature. Individual replicates are reported as circles.



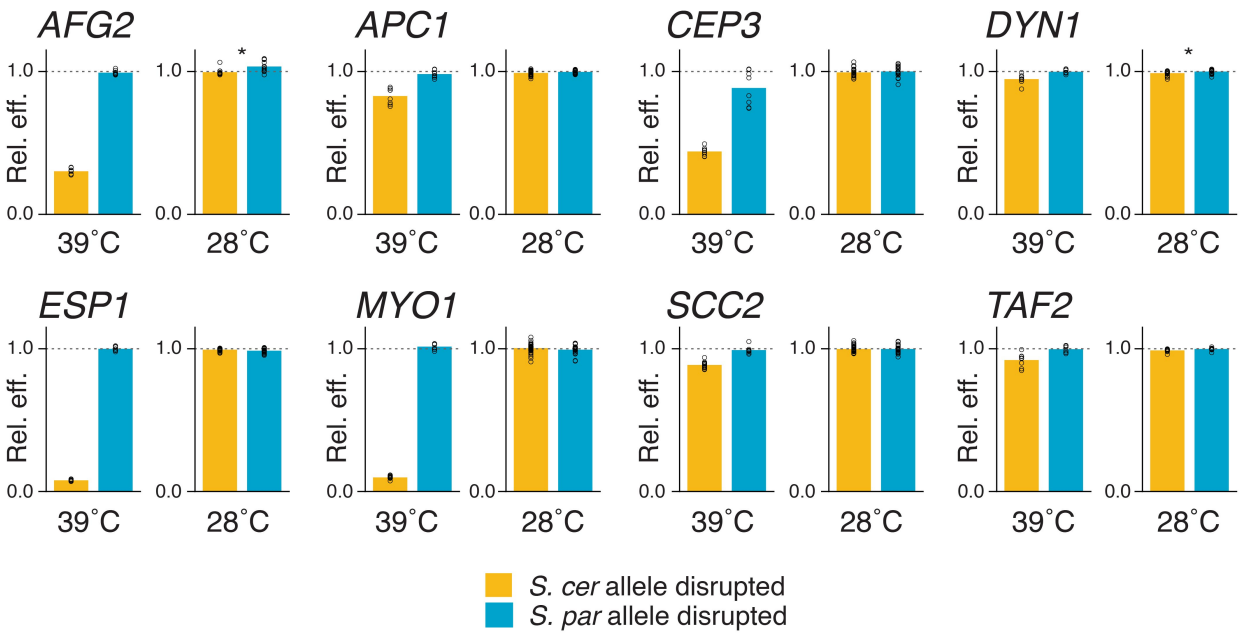
Supplementary Figure 4. Growth of the *S. cerevisiae* x *S. paradoxus* hybrid is between that of its purebred parents at 39°C.
Each bar reports the mean efficiency of the indicated strain (*n* = 4 cultures) after 24 hours of growth at 39°C; individual measurements are reported as circles. *, *p* = 0.035, two-sample, two-tailed *t*-test.



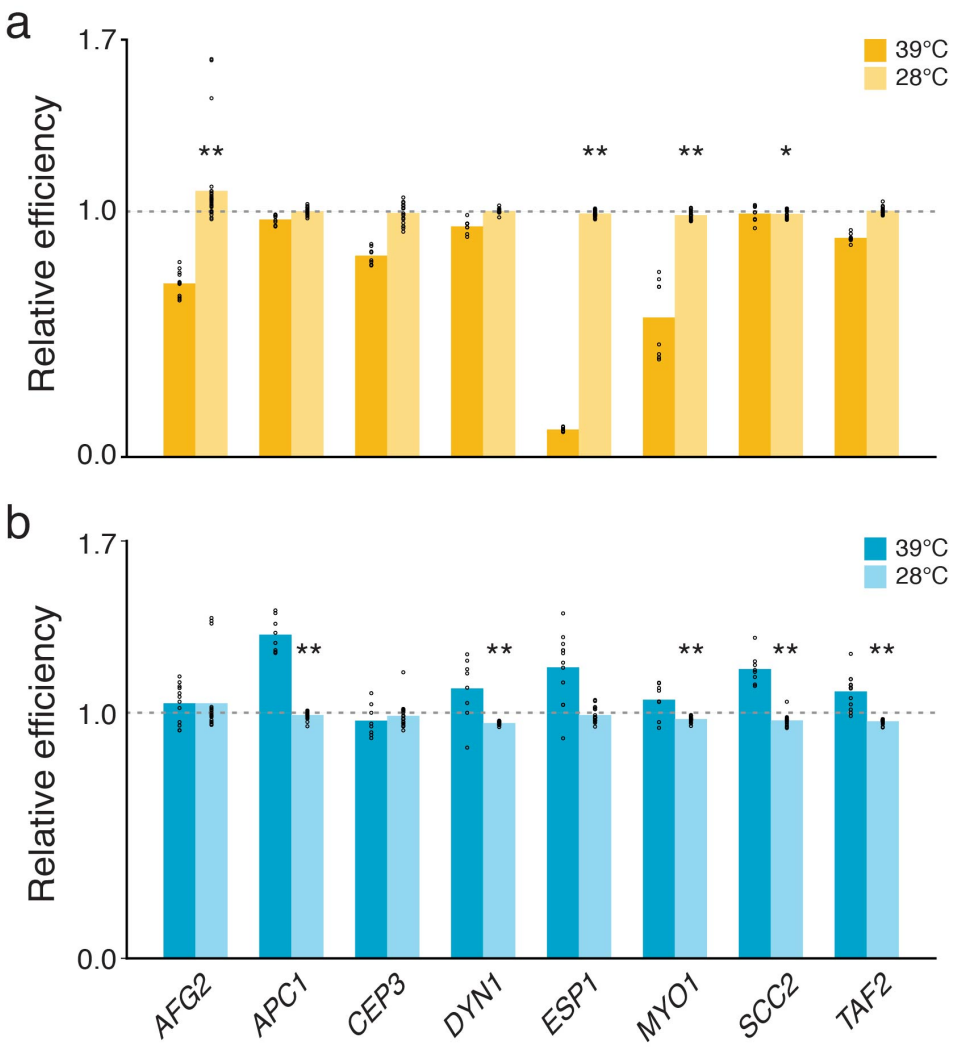
Supplementary Figure 5. Dependence of the RH-seq data set on cutoffs for read depth and transposon mutant coverage.
In each panel, the shading of a given point reports the size of the analyzed RH-seq data set or the set of mapped gene hits, upon filtering for the indicated depth and coverage attributes using cutoff values reported on the axes. **a**, The *x*-axis reports the average number of sequencing reads mapping to a given transposon insertion in either the 28°C or 39°C selection, as a minimum level above which the insertion was retained for analysis. The *y*-axis reports the coefficient of variation of read abundances between biological replicates for a given transposon insertion, as a maximum level below which the insertion was retained for analysis. The *z*-axis reports the number of transposon insertions per allele, as a minimum above which the gene was retained for analysis. Shading reports the number of insertions retained for analysis in the indicated cutoff scheme. **b**, Data and symbols are as in **a**, except that shading reports the number of genes retained for analysis in the indicated cutoff scheme. **c**, Data and symbols are as in **a**, except that shading reports the number of genes that scored below a *p*-value corresponding to a false-discovery rate (FDR) of 0.01 in the reciprocal hemizygosity test using the indicated cutoff scheme. Arrows indicate the set of cutoff values used in this study, which yielded a data set of 110,678 usable insertions across 3416 analyzable genes, 8 of which scored below FDR = 0.01 in the reciprocal hemizygosity test.


a
telomere repeats
Ty elements
rDNA
no genome sequence
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
log(10) reads/insert
insert position on chromosome (X100,000)
b
telomere repeats
Ty elements
rDNA
no genome sequence
log(10) reads/insert
insert position on chromosome (X100,000)

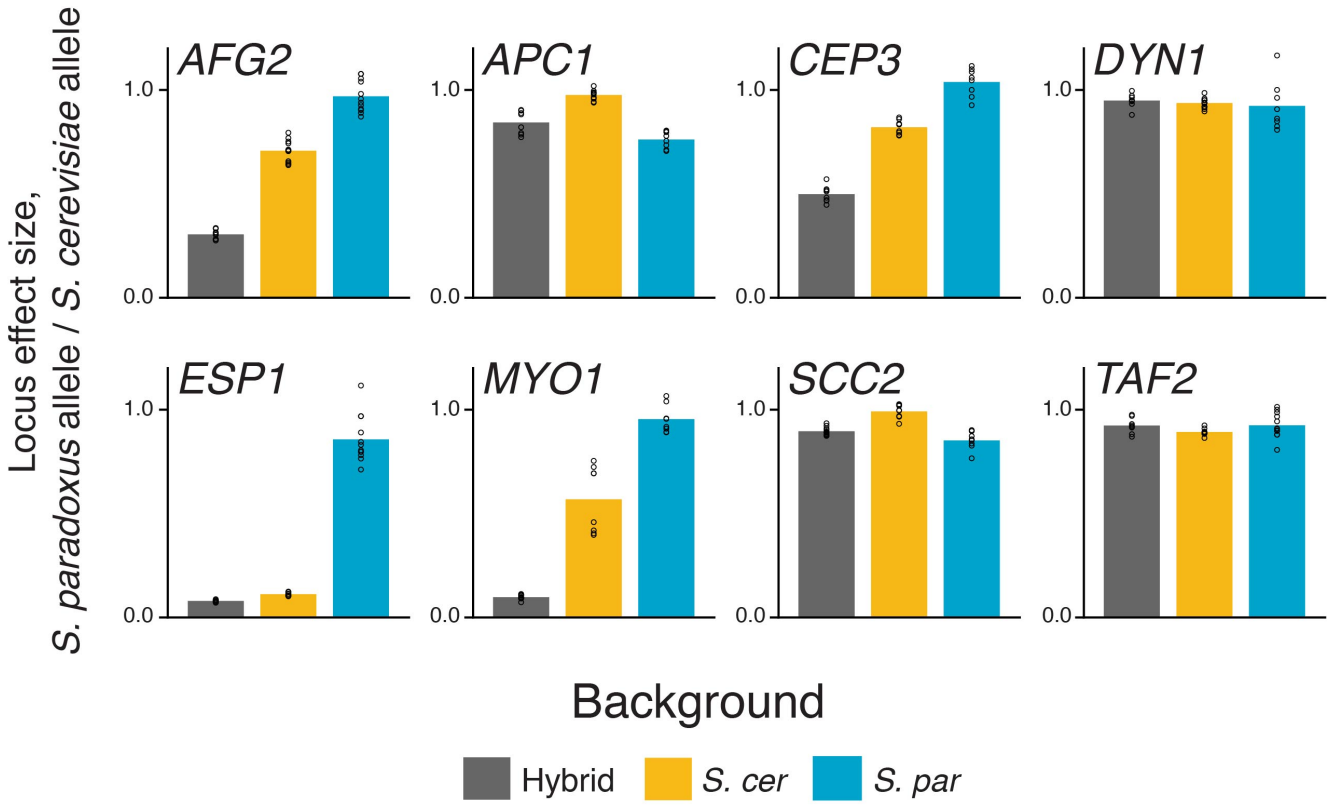
Supplementary Figure 6. RH-seq transposon coverage across the genome.
a, Each panel reports sites in which the PiggyBac transposon inserted in the indicated *S. cerevisiae* DBVPG1373 chromosome in clones of the *S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1 hybrid, as mapped from a pool of such clones by RH-seq. Each point reports one insertion; the *x*-axis reports the chromosomal position of a given insertion site, and the *y*-axis reports the raw number of sequencing reads mapped to that site. Colored tick marks along the bottom of each panel report genomic features that prohibited the mapping of reads. Read counts are from a representative RH-seq library after seven generations of culture at 39°C, reflecting the abundance in the pool of the respective hemizygote clone harboring the insertion. **b**, Data are as in **a**, except that shown are results from transposon insertions along *S. paradoxus* Z1 chromosomes in the *S. cerevisiae* x *S. paradoxus* hybrid.



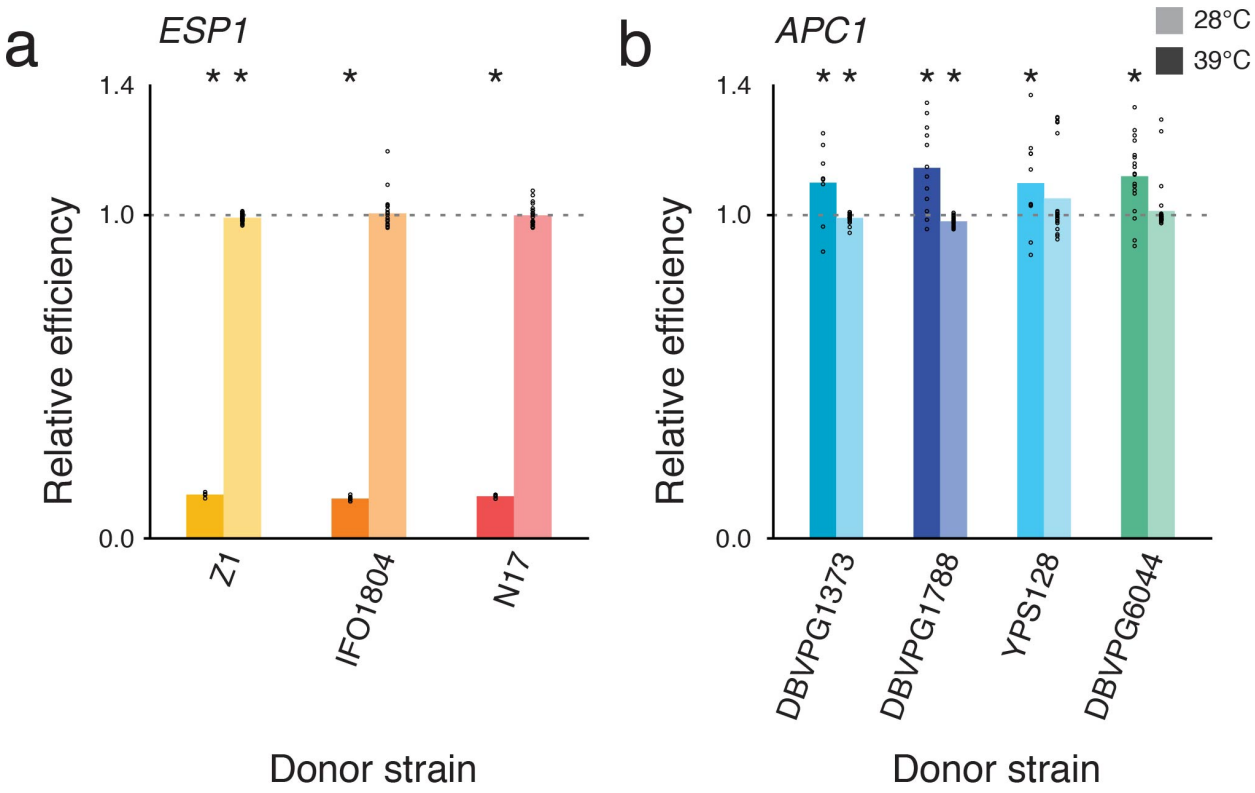
Supplementary Figure 7. Variation at RH-seq hit loci has little impact on growth at 28°C in the background of the interspecific hybrid.
Each panel reports growth efficiency measurements of targeted-deletion reciprocal hemizygotes at the indicated RH-seq hit locus. In a given panel, the left-hand pair of bars reports relative efficiencies of targeted-deletion hemizygotes after culture at 39°C, from Figure 2b of the main text. In the right-hand pair of bars, each bar reports the mean growth efficiency (n = 12-36 cultures) after culture at 28°C of a targeted-deletion hemizygote in the indicated species' allele, normalized by the analogous quantity for the wild-type hybrid parent; individual measurements are reported as circles. Statistical analyses of 39°C efficiency data are reported in Figure 2; *, $p \leq 0.05$, in a two-sample, two-tailed t-test for a difference in efficiency between the indicated hemizygotes at 28°C.



Supplementary Figure 8. Variation at RH-seq hit loci has little impact on growth at 28°C in the background of the purebred species.
a, Each pair of bars reports measurements of growth efficiency of an *S. cerevisiae* DBVPG1373 strain harboring the *S. paradoxus* Z1 allele at the indicated RH-seq hit locus, relative to the analogous quantity for wild-type *S. cerevisiae* DBVPG1373. The dark-shaded bar reports mean relative efficiency of the allele-replacement strain after culture at 39°C, from Figure 3 of the main text. The light-shaded bar reports mean growth efficiency (n = 20-33 cultures) of the allele replacement strain after culture at 28°C, relative to the analogous quantity for wild-type *S. cerevisiae* DBVPG1373; individual measurements are reported as circles. Statistical analyses of 39°C efficiency data are reported in Figure 3; *, $p \leq 0.05$ and **, $p \leq 0.01$, in a one-sample, two-tailed t-test for a difference in efficiency at 28°C between the indicated allele-replacement strain and the wild-type *S. cerevisiae* DBVPG1373. **b**, Data and symbols are as in **a**, except that each bar reports results from a *S. paradoxus* Z1 strain harboring the *S. cerevisiae* DBVPG1373 allele at the indicated locus, relative to wild-type *S. paradoxus* Z1.



Supplementary Figure 9. Effect sizes of thermotolerance loci depend on genetic background.
Each panel reports a comparison of the impact on thermotolerance of allelic variation at the indicated gene, in the indicated diploid backgrounds. In a given panel, the grey bar reports the mean growth efficiency at 39°C of a hybrid strain harboring a wild-type copy of the allele from *S. paradoxus* Z1 of the focal gene and a full deletion of the allele from *S. cerevisiae* DBVPG1373, normalized by the analogous quantity measured in a hybrid with a wild-type *S. cerevisiae* allele and a deletion in the *S. paradoxus* allele, from the insets of Figure 2 of the main text. The orange bar reports the mean growth efficiency at 39°C of a strain of the *S. cerevisiae* DBVPG1373 background harboring the allele from *S. paradoxus* Z1 of the focal gene, normalized by the analogous quantity measured in wild-type *S. cerevisiae* DBVPG1373, from Figure 3a of the main text. The blue bar reports the mean growth efficiency at 39°C of wild-type *S. paradoxus* Z1, normalized by the analogous quantity measured in a strain of *S. paradoxus* Z1 harboring the allele from *S. cerevisiae* DBVPG1373 of the focal gene, from Figure 3b of the main text. Individual measurements are reported as circles. Sample numbers for the grey, orange and blue bars are as reported in Figure 2 insets, Figure 3a and Figure 3b, respectively.



c

	mean, RH-seq hits	mean, genome	p
S.cer branch length	0.3026	0.2448	0.01
D_{xy} /length	0.1158	0.0993	0.04

Supplementary Figure 10. At RH-seq hit loci, the effect of allelic variation is conserved across a given species, and sequence divergence from *S. paradoxus* is a common feature of *S. cerevisiae* strains.
a, Each pair of bars reports growth efficiency of an *S. cerevisiae* DBVPG1373 strain harboring the allele of *ESP1* from the indicated strain of *S. paradoxus*, relative to the analogous quantity for wild-type *S. cerevisiae* DBVPG1373; heights of the dark and light bars report mean relative efficiency at 39°C (n = 4-18 cultures) and 28°C (n = 22-33 cultures), respectively. Individual measurements are reported as circles. **b**, Data and symbols are as in **a**, except that each bar reports results from a *S. paradoxus* Z1 strain harboring the allele of *APC1* from the indicated strain of *S. cerevisiae*, relative to wild-type *S. paradoxus* Z1. *, $p \leq 0.034$ in a one-sample, one-tailed t-test (39°C) or a one-sample, two-tailed t-test (28°C) for a difference in efficiency between the indicated allele-replacement strain and 1. Provenance of each strain is as follows: Z1, oak bark, UK; N17, oak exudate, Russia; IFO1804, oak bark, Japan; DBVPG1373, soil, Netherlands; DBVPG1788, soil, Finland; YPS128, soil, USA; DBVPG6044, bili wine, West Africa. **c,** Each row reports a comparison of the sequences of RH-seq hit loci against a genomic null. *S. cer* branch length, number of sequence substitutions along

the lineage leading to *S. cerevisiae,* in a phylogenetic tree inferred from *Saccharomyces* species type strains. D_{xy}/length, average number of differences between the *S. paradoxus* type strain and a strain randomly chosen from the *S. cerevisiae* wine/European population, normalized by gene length. The first and second columns report the average of the indicated statistic across the eight RH-seq hit loci and across sets of eight loci randomly resampled from the genome, respectively. The third column reports the empirical *p*-value from a test for an elevated value of the indicated statistic relative to the resampling null.

Supplementary Tables

Species	Strain name	Background	Genotype/details	Source	Allele donor strain
A. Wild-type diploid and hybrid parent strains					
S. cer	DBVPG1373		soil isolate from Netherlands	SGRP	
S. cer	YPS128		soil isolate from USA	SGRP	
S. cer	DBVPG1788		soil isolate from Finland	SGRP	
S. par	Z1		tree bark isolate from UK	SGRP	
S. par	DBVPG6304		isolate from *Drosophila* from USA	SGRP	
S. par	IFO1804		tree bark isolate from Japan	SGRP	
S. cer	JR501	DBVPG1373	ΔURA3::HygMX/ΔURA3::HygMX	this study	
S. par	JR499	Z1	ΔURA3::HygMX/ΔURA3::HygMX	this study	
S. cer x S. par hybrid	JR507	JR501x JR499	ΔURA3::HygMX/ΔURA3::HygMX, each S. par and S. cer allele KO'd with HygMX. This is the starting strain for PiggyBac mutagenesis.	this study	
S.cer	DBVPG6044		bili wine isolate from West Africa	SGRP	
S.par	N17		oak exudate isolate from Russia	SGRP	
B. Targeted-deletion reciprocal hemizygote strains					
S. cer x S. par hybrid	CW38	JR507	ΔscAFG2::KanMX/spAFG2	this study	
S. cer x S. par hybrid	CW39	JR507	ΔscAFG2::KanMX/spAFG2	this study	
S. cer x S. par hybrid	CW40	JR507	scAFG2/ΔspAFG2::KanMX	this study	
S. cer x S. par hybrid	CW41	JR507	scAFG2/ΔspAFG2::KanMX	this study	
S. cer x S. par hybrid	CW56	JR507	ΔscCEP3::KanMX/spCEP3	this study	
S. cer x S. par hybrid	CW57	JR507	ΔscCEP3::KanMX/spCEP3	this study	
S. cer x S. par hybrid	CW52	JR507	scCEP3/ΔspCEP3::KanMX	this study	
S. cer x S. par hybrid	CW53	JR507	scCEP3/ΔspCEP3::KanMX	this study	
S. cer x S. par hybrid	CW58	JR507	ΔscMYO1::KanMX/spMYO1	this study	
S. cer x S. par hybrid	CW59	JR507	ΔscMYO1::KanMX/spMYO1	this study	
S. cer x S. par hybrid	CW85	JR507	ΔscMYO1::KanMX/spMYO1	this study	
S. cer x S. par hybrid	CW60	JR507	scMYO1/ΔspMYO1::KanMX	this study	
S. cer x S. par hybrid	CW61	JR507	scMYO1/ΔspMYO1::KanMX	this study	
S. cer x S. par hybrid	CW80	JR507	ΔscSCC2::KanMX/spSCC2	this study	
S. cer x S. par hybrid	CW81	JR507	ΔscSCC2::KanMX/spSCC2	this study	
S. cer x S. par hybrid	CW82	JR507	ΔscSCC2::KanMX/spSCC2	this study	
S. cer x S. par hybrid	CW83	JR507	scSCC2/ΔspSCC2::KanMX	this study	
S. cer x S. par hybrid	CW84	JR507	scSCC2/ΔspSCC2::KanMX	this study	
S. cer x S. par hybrid	CW116	JR507	ΔscESP1::KanMX/spESP1	this study	
S. cer x S. par hybrid	CW122	JR507	ΔscESP1::KanMX/spESP1	this study	
S. cer x S. par hybrid	CW114	JR507	scESP1/ΔspESP1::KanMX	this study	
S. cer x S. par hybrid	CW193	JR507	scESP1/ΔspESP1::KanMX	this study	
S. cer x S. par hybrid	CW131	JR507	ΔscAPC1::KanMX/spAPC1	this study	

S. cer x S. par hybrid	CW160	JR507	ΔscAPC1::KanMX/spAPC1	this study	
S. cer x S. par hybrid	CW185	JR507	scAPC1/ΔspAPC1::KanMX	this study	
S. cer x S. par hybrid	CW189	JR507	scAPC1/ΔspAPC1::KanMX	this study	
S. cer x S. par hybrid	CW172	JR507	ΔscDYN1::KanMX/spDYN1	this study	
S. cer x S. par hybrid	CW173	JR507	ΔscDYN1::KanMX/spDYN1	this study	
S. cer x S. par hybrid	CW112	JR507	scDYN1/ΔspDYN1::KanMX	this study	
S. cer x S. par hybrid	CW113	JR507	scDYN1/ΔspDYN1::KanMX	this study	
S. cer x S. par hybrid	CW213	JR507	ΔscTAF2:NatMX/spTAF2	this study	
S. cer x S. par hybrid	CW214	JR507	ΔscTAF2:NatMX/spTAF2	this study	
S. cer x S. par hybrid	CW221	JR507	scTAF2/ΔspTAF2::NatMX	this study	
S. cer x S. par hybrid	CW222	JR507	scTAF2/ΔspTAF2::NatMX	this study	
C. Allele replacement strains					
S. cer	CW64	DBVPG1373	ΔAFG2(-343 to 2390) :: spAFG2(-365 to 2391)	this study	Z1
S. cer	CW65	DBVPG1373	ΔAFG2(-342 to 2391) :: spAFG2(-365 to 2391)	this study	Z1
S. cer	CW66	DBVPG1373	ΔAFG2(-342 to 2391) :: spAFG2(-365 to 2391)	this study	Z1
S. cer	CW73	DBVPG1373	ΔCEP3(-236 to 1831) :: spCEP3(-229 to 1831)	this study	Z1
S. cer	CW74	DBVPG1373	ΔCEP3(-236 to 1840) :: spCEP3(-229 to 1840)	this study	Z1
S. cer	CW104	DBVPG1373	ΔMYO1(-978 to 5848) :: spMYO1(-983 to 5849)	this study	Z1
S. cer	CW105	DBVPG1373	ΔMYO1(-978 to 5848) :: spMYO1(-983 to 5849)	this study	Z1
S. cer	CW98	DBVPG1373	ΔESP1(-384 to 4934) :: spESP1(-361 to 4933)	this study	Z1
S. cer	CW100	DBVPG1373	ΔESP1(-384 to 4934) :: spESP1(-361 to 4933)	this study	Z1
S. cer	CW109	DBVPG1373	ΔSCC2(-405 to 4543) :: spSCC2(-419 to 4543)	this study	Z1
S. cer	CW110	DBVPG1373	ΔSCC2(-377 to 4543) :: spSCC2(-388 to 4543)	this study	Z1
S. cer	CW115	DBVPG1373	ΔAPC1(-525 to 5270) :: spAPC1(-518 to 5270)	this study	Z1
S. cer	CW210	DBVPG1373	ΔAPC1(1626 to 5270) :: spAPC1(1623 to 5270)	this study	Z1
S. cer	CW254	DBVPG1373	ΔDYN1(-284 to 12340) :: spDYN1(-272 to 12340)	this study	Z1
S. cer	CW255	DBVPG1373	ΔDYN1(-284 to 12340) :: spDYN1(-272 to 12340)	this study	Z1
S. cer	CW174	DBVPG1373	ΔTAF2(-831 to 4328) :: spTAF2(-864 to 4319)	this study	Z1
S. cer	CW202	DBVPG1373	ΔTAF2(-831 to 4317) :: spTAF2(-864 to 4308)	this study	Z1
S. par	CW67	Z1	ΔAFG2(-293 to 2391) :: scAFG2(-270 to 2391)	this study	DBVPG1373
S. par	CW69	Z1	ΔAFG2(-293 to 2391) :: scAFG2(-270 to 2391)	this study	DBVPG1373
S. par	CW63	Z1	ΔCEP3(-187 to 1816) :: scCEP3(-194 to 1816)	this study	DBVPG1373
S. par	CW77	Z1	ΔCEP3(-187 to 1816) :: scCEP3(-194 to 1816)	this study	DBVPG1373
S. par	CW124	Z1	ΔMYO1(-907 to 5826) :: scMYO1(-903 to 5824)	this study	DBVPG1373
S. par	CW126	Z1	ΔMYO1(-907 to 5826) :: scMYO1(-903 to 5824)	this study	DBVPG1373
S. par	CW141	Z1	ΔESP1(-418 to 4962) :: scESP1(-439 to 4963)	this study	DBVPG1373
S. par	CW142	Z1	ΔESP1(-418 to 4951) :: scESP1(-439 to 4952)	this study	DBVPG1373
S. par	CW165	Z1	ΔSCC2(-456 to 4556) :: scSCC2(-441 to 4556)	this study	DBVPG1373
S. par	CW166	Z1	ΔSCC2(-456 to 4556) :: scSCC2(-441 to 4556)	this study	DBVPG1373
S. par	CW132	Z1	ΔAPC1(-538 to 5280) :: scAPC1(-545 to 5280)	this study	DBVPG1373

S. par	CW133	Z1	ΔAPC1(-542 to 5275) :: scAPC1(-549 to 5275)	this study	DBVPG1373
S. par	CW169	Z1	ΔTAF2(-806 to 4348) :: scTAF2(-774 to 4357)	this study	DBVPG1373
S. par	CW170	Z1	ΔTAF2(-806 to 4348) :: scTAF2(-774 to 4357)	this study	DBVPG1373
S. par	CW223	Z1	ΔDYN1(-231 to 12377) :: scDYN1(-243 to 12377)	this study	DBVPG1373
S. par	CW224	Z1	ΔDYN1(-231 to 12377) :: scDYN1(-243 to 12377)	this study	DBVPG1373
S. cer	CW284	DBVPG1373	ΔESP1(-384 to 4934) :: spESP1(-361 to 4934)	this study	IFO1804
S. cer	CW288	DBVPG1373	ΔESP1(-384 to 4934) :: spESP1(-361 to 4934)	this study	IFO1804
S. cer	CW287	DBVPG1373	ΔESP1(-384 to 4934) :: spESP1(-361 to 4933)	this study	N17
S. cer	CW324	DBVPG1373	ΔESP1(-384 to 4934) :: spESP1(-361 to 4933)	this study	N17
S.par	CW300	Z1	ΔAPC1(-547 to 5280) :: scAPC1(-554 to 5280)	this study	YPS128
S.par	CW322	Z1	ΔAPC1(-547 to 5280) :: scAPC1(-554 to 5280)	this study	YPS128
S.par	CW307	Z1	ΔAPC1(-547 to 5280) :: scAPC1(-554 to 5280)	this study	DBVPG1788
S.par	CW308	Z1	ΔAPC1(-547 to 5280) :: scAPC1(-554 to 5280)	this study	DBVPG1788
S.par	CW309	Z1	ΔAPC1(-547 to 5280) :: scAPC1(-554 to 5280)	this study	DBVPG6044
S.par	CW310	Z1	ΔAPC1(-547 to 5280) :: scAPC1(-554 to 5280)	this study	DBVPG6044
S.par	CW311	Z1	ΔAPC1(-547 to 5280) :: scAPC1(-554 to 5280)	this study	DBVPG6044

Supplementary Table 1. Strains used in this study. a, Wild-type diploid strains, including those used as parents of the *S. cerevisiae* x *S. paradoxus* hybrid and of allele-replacement transgenesis; SGRP, the Saccharomyces Genome Resequencing Project, version 2. **b**, Hemizygotes in the *S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1 diploid hybrid constructed by targeted deletion of a given species' allele of the indicated gene with the KanMX or NatMX cassette. ΔscYFG::KanMX/spYFG signifies that the *S. cerevisiae* DBVPG1373 allele of YFG was knocked out and the *S. paradoxus* Z1 allele of YFG is intact; strains with the DBVPG1373 allele intact and the Z1 allele knocked out are represented analogously. **c**, Allele replacement strains in *S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1 diploid homozygote backgrounds. In genotype notes, *e.g.* in an *S. paradoxus* background, ΔYFG(-X to +Y)::scYFG(-Z to +W) indicates that in *S. paradoxus* Z1, bases -X to +Y from gene YFG have been removed and replaced by bases -Z to +W of the allele of YFG from the indicated *S. cerevisiae* strain. Positive coordinates count in the 5' to 3' direction from the start codon (+1 corresponds to the A in the ATG), and negative coordinates count in the 3' to 5' direction from the start codon (-1 corresponds to the base directly 5' of the ATG). In cases where the replacement extended into a region of 100% conservation between species, the position of the last divergent nucleotide is shown.

References

1 Orr, H.A. The genetics of species differences. *Trends Ecol Evol* **16**, 343-350 (2001).

2 Flint, J. & Mott, R. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat Rev Genet* **2**, 437-445, doi:10.1038/35076585 (2001).

3 Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature*, doi:10.1038/nature24287 (2017).

4 Savolainen, O., Lascoux, M. & Merila, J. Ecological genomics of local adaptation. *Nat Rev Genet* **14**, 807-820, doi:10.1038/nrg3522 (2013).

5 Nadeau, N. J. & Jiggins, C. D. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet* **26**, 484-492, doi:10.1016/j.tig.2010.08.004 (2010).

6 Wray, G. A. Genomics and the Evolution of Phenotypic Traits. *Annual Review of Ecology, Evolution, and Systematics* **44**, 51-72 (2013).

7 Masly, J. P. & Presgraves, D. C. High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS biology* **5**, e243, doi:10.1371/journal.pbio.0050243 (2007).

8 Greig, D. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS genetics* **3**, e21, doi:10.1371/journal.pgen.0030021 (2007).

9 Eshed, Y. & Zamir, D. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147-1162 (1995).

10 Lazzarano, S. *et al.* Genetic mapping of species differences via in vitro crosses in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **115**, 3680-3685, doi:10.1073/pnas.1717474115 (2018).

11 Roop, J. I., Chang, K. C. & Brem, R. B. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature* **530**, 336-339, doi:10.1038/nature16938 (2016).

12 Steinmetz, L. M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326-330, doi:10.1038/416326a (2002).

13 Stern, D. L. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends Genet* **30**, 547-554, doi:10.1016/j.tig.2014.09.006 (2014).

14 Goncalves, P., Valerio, E., Correia, C., de Almeida, J. M. & Sampaio, J. P. Evidence for divergent evolution of growth temperature preference in sympatric Saccharomyces species. *PLoS One* **6**, e20739, doi:10.1371/journal.pone.0020739 (2011).

15 Salvado, Z. *et al.* Temperature adaptation markedly determines evolution within the genus Saccharomyces. *Appl Environ Microbiol* **77**, 2292-2302, doi:10.1128/AEM.01861-10 (2011).

16 Sweeney, J. Y., Kuehne, H. A. & Sniegowski, P. D. Sympatric natural Saccharomyces cerevisiae and S. paradoxus populations have different thermal growth profiles. *FEMS Yeast Res* **4**, 521-525 (2004).

17 Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Bethesda)* **1**, 11-25, doi:10.1534/g3.111.000273 (2011).
18 Hartwell, L. H. Saccharomyces cerevisiae cell cycle. *Bacteriol Rev* **38**, 164-198 (1974).
19 Mitra, R., Fain-Thornton, J. & Craig, N. L. piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J* **27**, 1097-1109, doi:10.1038/emboj.2008.41 (2008).
20 van Opijnen, T., Lazinski, D. W. & Camilli, A. Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. *Curr Protoc Mol Biol* **106**, 7 16 11-24, doi:10.1002/0471142727.mb0716s106 (2014).
21 Parts, L. *et al.* Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res* **21**, 1131-1138, doi:10.1101/gr.116731.110 (2011).
22 Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).
23 Leuenberger, P. *et al.* Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, doi:10.1126/science.aai7825 (2017).
24 Wilkening, S. *et al.* An evaluation of high-throughput approaches to QTL mapping in Saccharomyces cerevisiae. *Genetics* **196**, 853-865, doi:10.1534/genetics.113.160291 (2014).
25 Neurological Disorders: Public Health Challenges. 232 (World Health Organization, 2006).
26 Sinha, H., Nicholson, B. P., Steinmetz, L. M. & McCusker, J. H. Complex genetic interactions in a quantitative trait locus. *PLoS Genet* **2**, e13, doi:10.1371/journal.pgen.0020013 (2006).
27 Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**, 2519-2524 (1996).
28 Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**, e00306-00315, doi:10.1128/mBio.00306-15 (2015).
29 Skelly, D. A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* **23**, 1496-1504, doi:10.1101/gr.155762.113 (2013).
30 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
31 Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337-341, doi:10.1038/nature07743 (2009).
32 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
33 Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, 1987).

34 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
35 Loytynoja, A. & Goldman, N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**, 579, doi:10.1186/1471-2105-11-579 (2010).
36 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).
37 Roop, J. I. & Brem, R. B. Rare variants in hypermutable genes underlie common morphology and growth traits in wild Saccharomyces paradoxus. *Genetics* **195**, 513-525, doi:10.1534/genetics.113.155341 (2013).
38 Maurer, M. J. *et al.* Quantitative Trait Loci (QTL)-Guided Metabolic Engineering of a Complex Trait. *ACS Synth Biol* **6**, 566-581, doi:10.1021/acssynbio.6b00264 (2017).
39 Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol* **4**, 975-986, doi:10.1021/sb500366v (2015).

Chapter 4

S. cerevisiae exhibits a derived program of sensitivity to cold and a yeast microtubule poison

The following work was done in partnership with an undergraduate in the lab, Anna Flury.

Abstract

Although trait differences between species often evolve through genetic drift, environmental niche differences can also drive the evolution of distinct traits through the processes of natural selection. It is often these adaptive traits that are the most exciting for evolutionary geneticists to study. We have uncovered a phenotypic difference in resistance to a microtubule poison (benomyl) between species of *Saccharomyces* yeasts separated by ~5 million years of evolution. Benomyl resistance is dependent on temperature and physical environment, hinting at the potential environmental requirements that selected for it. And although this difference is generally consistent amongst strains of *S. cerevisiae* and *S. paradoxus*, and therefore likely a result of selective processes, we show that North American *S. cerevisiae* strains retain the putatively ancestral benomyl tolerance phenotype. Finally, we show that resistance to benomyl is correlated to cold resistance in a dataset of almost 1,000 geographically diverse *S. cerevisiae* strains. Our results suggest a model tying together the cold sensitivity of *S. cerevisiae* to microtubule stability, potentially resolving the question of the genetic basis of the latter, which has been observed for decades but was never mechanistically understood.

Introduction

Why study a microtubule poison?

As in Chapter 3, the following work was motivated by a desire to understand the genetic basis of a trait that arose long ago, has fixed in one species relative to another, and has been shaped by natural selection. Such traits are actually somewhat difficult to come by. What we more often observe is variation between the strains or individuals of a given species of similar magnitude to variation between species. Therefore, by random chance, one can end up selecting individual tester strains that make it appear as if there is a true species-specific difference in a focal trait, when in fact it is an artifact of ascertainment bias. In such a case it is difficult to assess whether genetic drift or selection due to a true environmental/niche difference has caused the observed phenotypic change. Conversely, when all of the strains of a given species have a specific phenotype, which differs from all other strains of another species, it is much more parsimonious to hypothesize that a difference in selective pressures between the species is or was at play. As such, we are always on the lookout for traits that seem to truly represent species-specific differences amongst *Saccharomyces*.

Over the course of searching for phenotypic differences between *S. cerevisiae* and *S. paradoxus*, I came across an unexpected result. Data from our work on thermotolerance had led us to hypothesize that *S. cerevisiae* may have chromosome segregation machinery that, at high temperature, is particularly robust compared to that of *S. paradoxus.* It is generally accepted that susceptibility to microtubule poisons, e.g.

the drugs benomyl or nocodazole, indicates that something is going wrong with chromosome segregation as the cell divides, and this has been used in the past as a tool to study the cell cycle[1,2]. On this basis, I originally hypothesized that *S. cerevisiae* would be more resistant to benomyl than *S. paradoxus*. It turned out that we were incorrect in this hypothesis; as is often the case in science, an unexpected finding turned into its own interesting project, which we believe sheds new light on the ecological genetics of *Saccharomyces*.

Results

An interspecific difference in benomyl tolerance

Despite my original expectation that *S. paradoxus* would be more susceptible to benomyl compared to *S. cerevisiae*, the reverse is true: at 23°C, the Z1 strain of *S. paradoxus,* an oak tree isolate, grows well at concentrations of benomyl that greatly inhibit the DBVPG1373 strain of *S. cerevisiae,* collected from soil in the Netherlands (Figure 1). We were able to replicate this finding in a quantitative assay through competitive growth on agar plates and qPCR to assess abundance of each species (Supplementary Figure 1). By contrast, at temperatures at and above 28°C, *S. cerevisiae* DBVPG1373 can grow well on benomyl plates (Figure 2, bottom three panels). At very high temperatures (37°C), the interspecific difference between *S. cerevisiae* and *S. paradoxus* appears to be reversed; this is likely due to the confounding inhibitory effect of high temperature on *S. paradoxus* Z1 growth, as its growth on control DMSO plates is also lowered at this temperature (Figure 2, bottommost panel). And in further contrast to the result from Figure 1 that benomyl differentially inhibits growth of *S. cerevisiae* DBVPG1373 and *S. paradoxus* Z1 on solid YPD-agar plates, this difference disappears when growing these strains in liquid YPD containing DMSO or benomyl; *S. paradoxus* Z1 is no longer resistant to benomyl in liquid (Figure 3). Liquid-culture experiments done at lower benomyl concentrations and at different temperatures showed similar results (data not shown). We conclude that *S. paradoxus* Z1 is strongly benomyl resistant relative to *S. cerevisiae* DBVPG1373, but that this striking difference depends on temperature and physical environment.

Lack of resistance to benomyl is a derived trait in most S. cerevisiae strains

We next asked whether benomyl resistance in *S. paradoxus* was newly evolved, i.e. a derived trait only present in this species, or an ancestral trait shared by other *Saccharomyces.* As shown in Figure 4, representative strains of other *Saccharomyces* species (*S. bayanus, S. mikatae, S. arboricolus* and *S. kudriavzevii*) all exhibit levels of benomyl resistance similar to that of *S. paradoxus* Z1, indicating that resistance to benomyl is the ancestral state for this phenotype. Along these lines, we also wondered whether loss of benomyl resistance is a trait consistent amongst *S. cerevisiae* strains from populations in different parts of the world, or if it is a loss unique to DBVPG1373 (conversely, we wondered whether all *S. paradoxus* strains still maintain benomyl resistance or if it is unique to Z1). We saw that all *S. paradoxus* strains we tested had similar levels of benomyl resistance (Figure 5a). In addition, we saw that almost all *S.*

cerevisiae strains had lost benomyl resistance (Figure 5b). Surprisingly, however, we noticed that one group of strains all still maintained benomyl resistance at levels similar to *S. paradoxus* – strains of *S. cerevisiae* collected from North America (e.g. YPS606 [Figure 5b] and other North American strains, data not shown). Together, these data suggest the ancestral state in the *Saccharomyces* clade was one of benomyl tolerance and that most, but not all, strains of *S. cerevisiae* have lost this character relatively recently.

Correlation of benomyl resistance to cold tolerance

If we seek to understand the evolutionary forces driving the putative maintenance of benomyl resistance in most *Saccharomyces* species, and loss of resistance in *S. cerevisiae,* we need to look beyond the drug itself, to which yeast is presumably not exposed in the wild. We noted that the molecular target of benomyl is microtubules, and microtubules are cold labile (see below for a more in-depth discussion). As such, we formulated a model in which benomyl tolerance could be a stand-in for cold tolerance, which is a much more ecologically relevant trait. To begin to test this notion, we used published measurements of growth on solid agar plates at 30°C of almost 1,000 strains from *S. cerevisiae* collected from around the globe[3]. We found that growth at either low (200 µg/mL) or high (500 µg/mL) concentrations of benomyl was correlated with growth at low temperatures (14°C), to different degrees (r-value = 0.50, *p*-value = $3.2e^{-63}$; r-value = 0.37, *p*-value = $1.3e^{-33}$, respectively). As a point of reference, we also calculated the correlation between low and high concentrations of benomyl, and found that to be the most highly correlated comparison (r-value = 0.74, *p*-value = $2.3e^{-166}$). We conclude that, in general, the most cold-tolerant *S. cerevisiae* strains also tend to be more resistant to benomyl.

Next, we asked if the relationship between benomyl and cold response would extend to a comparison across species. This prediction bore out in growth experiments at 4°C: *S. cerevisiae* DBVPG1373 grew very little on solid YPD agar plates after 1 month of incubation compared to *S. paradoxus* Z1 (Figure 6). As expected, the North American strains of *S. cerevisiae,* which our previous work had shown to be the most resistant with respect to benomyl, also showed improved cold tolerance at 4°C, although their growth did not match that of *S. paradoxus*.

Discussion

Benomyl tolerance and cold tolerance

In this work, we have shown that a derived cold- and benomyl- sensitive growth phenotype distinguishes *S. cerevisiae* from other species in its clade, with one exception: a few *S. cerevisiae* strains from one niche, that of North America, exhibit resistance to these treatments instead.

Although at first glance benomyl and cold stress may seem to have nothing in common, the classic cell biological literature suggests a strong mechanistic connection.

The molecular target of benomyl is microtubules, specifically the protein product of *TUB2,* beta-tubulin[4]. Microtubules in laboratory *S. cerevisiae* and other model systems function well at permissive temperatures but are cold labile[5]: both cytoplasmic and spindle microtubules dissociate in laboratory *S. cerevisiae* after being incubated at 4°C for 24 hours[6].

On the basis of these trends and our own data, we thus favour a model in which microtubules of most *S. cerevisiae* strains are sensitive to cold and benomyl, which governs the growth disadvantage of these isolates relative to other yeast species (and North American *S. cerevisiae*). Given the inferred historical origin of *S. cerevisiae* in warm locations of East Asia[3], it is tempting to speculate that relaxed selection on cold tolerance may have driven the accumulation of variants that compromise microtubule stability in this species, with later compensation in North American strains. Future work will test this notion, and the microtubule stability of strains and species across the clade.

Connection of benomyl resistance to freeze-thaw and sorbitol tolerance

Our studies to date leave open the genetic basis of the differences between yeast strains and species in cold and benomyl resistance. However, the variation we have seen among *S. cerevisiae* provides an intriguing clue, because it echoes almost perfectly a different phenotypic syndrome studied years ago by another group. Will et. al showed that North American *S. cerevisiae* strains have retained an ancestral freeze-thaw tolerance phenotype similar to *S. paradoxus*, whereas all other *S. cerevisiae* strains have lost freeze-thaw tolerance relatively recently[7]. They mapped the genetic basis of this difference between *S. cerevisiae* strains to two genes, *AQY1* and *AQY2,* which encode aquaporins (water channel membrane proteins). Not only did *AQY1* and *AQY2* mediate freeze-thaw tolerance, the authors also provided evidence to suggest that there is a trade off between freeze-thaw tolerance and osmotolerance (growth in high-sugar environments): *S. cerevisiae* strains with low freeze-thaw tolerance, lacking aquaporins, had relatively reduced growth in sorbitol upon the addition of aquaporins back into the strains, and strains with high freeze-thaw tolerance, with functioning aquaporins, had relatively higher growth in sorbitol upon deletion of the *AQY* genes. By examining the sequences of the *AQY* genes, these authors noticed that freeze-thaw sensitive *S. cerevisiae* strains harbour loss-of-function alleles at one or both *AQY* genes, resulting from a variety of different mutations (a combination of deletions and insertions, resulting in frame-shifts at different parts of the gene), depending on the strain. The authors concluded that there have been multiple, independent adaptive losses of aquaporins, resulting in improved osmotolerance, in most *S. cerevisiae* strains – except the North American strains. It is as yet undetermined what local, environmental niche differences distinguish the North American *S. cerevisiae* strains from other European or African *S. cerevisiae* strains, and how these unknown environment differences could have contributed to differential selective pressures across *S. cerevisiae* strains, or between *S. cerevisiae* and *S. paradoxus*.

Given the striking similarity between the published pattern of strains showing freeze-thaw tolerance on the one hand, and our observations of cold and benomyl

tolerance on the other (North American *S. cerevisiae* strains having a putatively ancestral phenotype with respect to both compared to most other *S. cerevisiae* strains), we consider it a very strong hypothesis that the genetic basis of the cold and benomyl tolerance phenotypes we have observed is also related to aquaporins. Our model is that ancestral, functional *AQY* alleles in North American *S. cerevisiae* and all other *Saccharomyces* species underlie resistance to benomyl and cold, and that loss-of-function alleles in most *S. cerevisiae* strains are responsible for their inability to grow under benomyl or cold stress. Further work will be necessary to directly test this notion, and, if it proves wrong, to dissect in an unbiased manner the variation in benomyl and cold resistance we see in this clade.

Materials and Methods

Strains and media

Strains used in this study are listed in Supplementary Table 1. Yeast peptone dextrose (YPD, 1% yeast extract [BD], 2% yeast peptone [BD], 2% D-glucose [Sigma]) was used as sterilized (YP dissolved and sterilized via autoclave, then mixed to 2% D-glucose with filter sterilized D-glucose) and used in liquid culture or mixed with 2% agar for use in solid agar plates. Benomyl (Spectrum Chemicals) was dissolved in dimethyl sulfoxide (DMSO) to a concentration of 10 mg/mL and stored at -20°C until use at the concentrations indicated in each figure. Control experiments were run with DMSO at the same concentration as existed in each benomyl experiment.

Growth assays

Solid

For growth assays of a given *Saccharomyces* strain on solid agar plates (Figures 1, 2, 4, 5, 6), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 28°C for 3 days. 1-2 colonies of each strain were inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm. In the case of where a single colony was struck, the single pre-culture from each strain was then split into two replicate pre-cultures and each again back-diluted into YPD in 1-inch diameter glass tubes with a target OD_{600} of 0.05; in cases where two colonies were struck, each replicate pre-culture was back-diluted into YPD in 1-inch diameter glass tubes with a target OD_{600} of 0.05. In Figures 4, 5 and 6, only a single non-split pre-culture was used (no replicates were carried out). In any such case, each back-diluted pre-culture was then grown at 28°C with shaking at 200rpm for 5.5-6 hours, and OD_{600} was measured again. Each replicate culture was normalized to the lowest OD_{600} out of all the replicate pre-cultures of a given day by back-dilution with extra liquid YPD. Two replicate 10-fold dilution series were made of each culture using YPD, up to a 1000-fold dilution; 3µL of each step of each dilution series, starting with undiluted on the outermost dot, was dotted onto a single YPD agar plate (with or without DMSO or benomyl, at the concentrations indicated in each figure; the control DMSO concentration was matched to the corresponding concentration of DMSO present in the comparing benomyl plate), with all dilution series of each replicate pre-culture of each strain being dotted onto the same plate for a given photo. Each plate was incubated at the temperature indicated in each figure until growth was observed, generally 2-3 days (for growth at 4°C, plates were incubated for ~1 month). The pipeline from inoculation off solid plates through pre-culture, back-dilution, and growth on solid we refer to as a day's growth experiment. 1-2 day's growth experiments were done for each figure.

Liquid

For growth measurements of a given *Saccharomyces* strain in liquid (Figure 3), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and

incubated at 28°C for 3 days. A single colony of each strain was each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate a pre-culture. Each was back-diluted into YPD at an OD_{600} of 0.05 and grown for an additional 5.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. Each pre-culture were then split into two and each again back-diluted into YPD, YPD + 0.5% DMSO or YPD+ 50 µg/mL benomyl in 1-inch diameter glass tubes with a target OD_{600} of 0.05; the actual OD_{600} of each was measured, after which it was grown at 23°C with shaking at 200rpm for 24 hours, and OD_{600} was measured again. The efficiency for each replicate was calculated as the difference between these final and initial OD_{600} values. The mean efficiency was calculated as the average between replicate cultures for each species and condition.

Competitive growth assay with quantitative PCR (qPCR)

For competitive growth measurements of *Saccharomyces* species on solid plates (Supplementary Figure 1), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 28°C for 3 days. 1 colony of each strain was inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm. Each was back-diluted into YPD at an OD_{600} of 0.05 and grown for an additional 5.5-6 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase, and OD_{600} was measured again. Each replicate culture was normalized to the lowest OD_{600} out of all the replicate pre-cultures of a given day by back-dilution with extra liquid YPD. For each competition, 100 µL of each of the two strains to be competed were mixed, and the entire 200 µL was plated onto a single agar plate (either YPD, YPD + 0.1% DMSO or YPD + 10 µg/mL benomyl). Plates were grown for ~48 hours at 23°C until a lawn formed, then the entire plate's worth of cells were scraped off of the agar into 45 mL of sterile water. 1 mL of cell suspension was removed, vortexed and genomic DNA was isolated from the pellet. For each such plate, qPCR reactions were set up in technical triplicate (three identical qPCR reactions for each sample and primer pair), using sets of allele-specific primers to either *S. paradoxus ACT1* (F: 5'-TTACGTCGGTGATGAAGCC-3', R: 5'-AATTGGAACAACGTGAGTG-3', or *S. cerevisiae ACT1* (F: 5'-CGTTCCAATTTACGCTGG-3', R: 5'-AGAAGATTGAGCAGCGGT-3'). All qPCR was done using the DyNAmo HS SYBR Green qPCR kit (Thermo Fisher Scientific) with ~100 ng of gDNA per reaction and a program of: 95°C – 15 minutes, 94°C – 10 seconds, 55°C – 30 seconds, 72°C – 30 seconds, x40 cycles, then melting curve analysis. Initially, a five-point standard curve was generated with 10-fold dilutions of either *S. cerevisiae* or *S. paradoxus* genomic DNA, for each respective primer pair, to calculate the effective efficiency, E, of each primer pair (calculated efficiencies: *S. cerevisiae:* 1.9163, *S. paradoxus:* 1.8633). For each competition in Supplementary Figure 1 (YPD, DMSO or benomyl), the C_t for each primer pair was measured in technical triplicate, and the DNA quantity for each of the triple replicates for each primer pair (D_{sc} or D_{sp}) was calculated as E^{-Ct}. The means of each of D_{sc} and D_{sp}, $\langle D_{sc} \rangle$ and $\langle D_{sp} \rangle$, respectively, were calculated for each competition from the triplicates. Thus, the final values were $\langle D_{sc} \rangle$ and $\langle D_{sp} \rangle$ for each competition done on YPD, YPD+DMSO or YPD+benomyl. For Supplementary Figure 1, we took the ratio of $\langle D_{sc} \rangle$ from YPD+benomyl to $\langle D_{sc} \rangle$ from YPD+DMSO, $\langle D_{sc} \rangle_{benomyl}$ /

$<D_{sc}>_{DMSO}$, as well as $<D_{sp}>$ from YPD+benomyl to $<D_{sp}>$ from YPD+DMSO, $<D_{sp}>_{benomyl} / <D_{sp}>_{DMSO}$, to control for any growth differences that existed without benomyl. Finally, we normalized each of these ratios by dividing by $<D_{sp}>_{benomyl} / <D_{sp}>_{DMSO}$, effectively setting the relative amount of *S. paradoxus* DNA in each competition to 1.0.

Correlation testing

Phenotype data on 971 geographically diverse *S. cerevisiae* strains was obtained from the Supplementary Data in [3]. A linear least-squares regression was calculated on data from the 200 µg/mL benomyl, 500 µg/mL benomyl and 14°C conditions using the scipy.stats “lingregress” function in Python. It is worth noting again that the data from this paper were collected at 30°C on solid YPD agar plates.

Figures

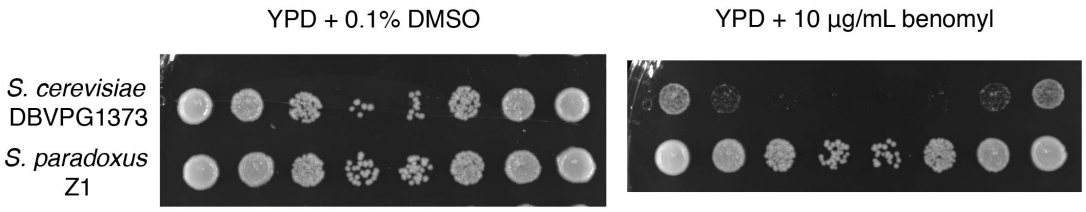


Figure 1. *S. paradoxus* Z1 has improved growth on solid media containing benomyl compared to *S. cerevisiae*. Each image represents a single YPD plate containing either 0.1% DMSO or 10 µg/mL benomyl, as indicated grown for 2 days at 23°C. Each dot is liquid culture from either *S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1 as indicated, starting from concentrated (outermost dots) to less concentrated (innermost dots), moving in 10-fold dilutions. Left and right dilution series represent technical replicates of a dilution series made from the same biological culture. Two biological cultures were grown and plated to the same plate on the same day; representative dilution series are shown.

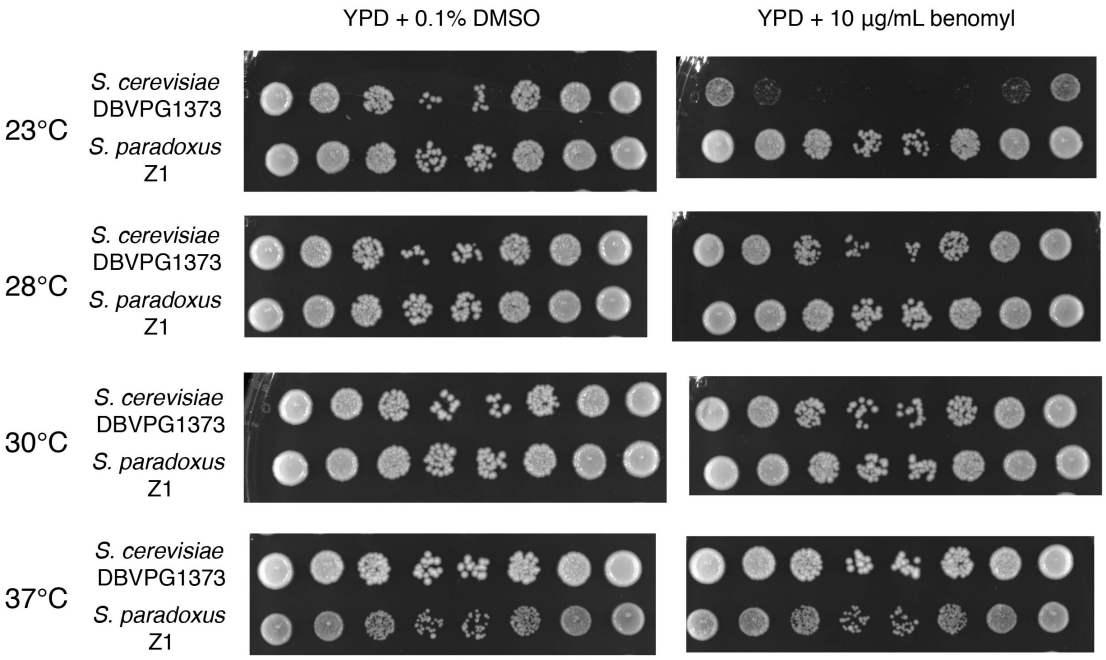


Figure 2. *S. cerevisiae* benomyl susceptibility is temperature dependent. Each image represents a single YPD plate containing either 0.1% DMSO or 10 µg/mL benomyl, as indicated grown for 2 days at the indicated temperature. Each dot is liquid culture from either *S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1 as indicated, starting from concentrated (outermost dots) to less concentrated (innermost dots), moving in 10-fold dilutions. Left and right dilution series represent technical replicates of a dilution series made from the same biological culture. Two biological cultures were grown and plated to the same plate on the same day; representative dilution series are shown.

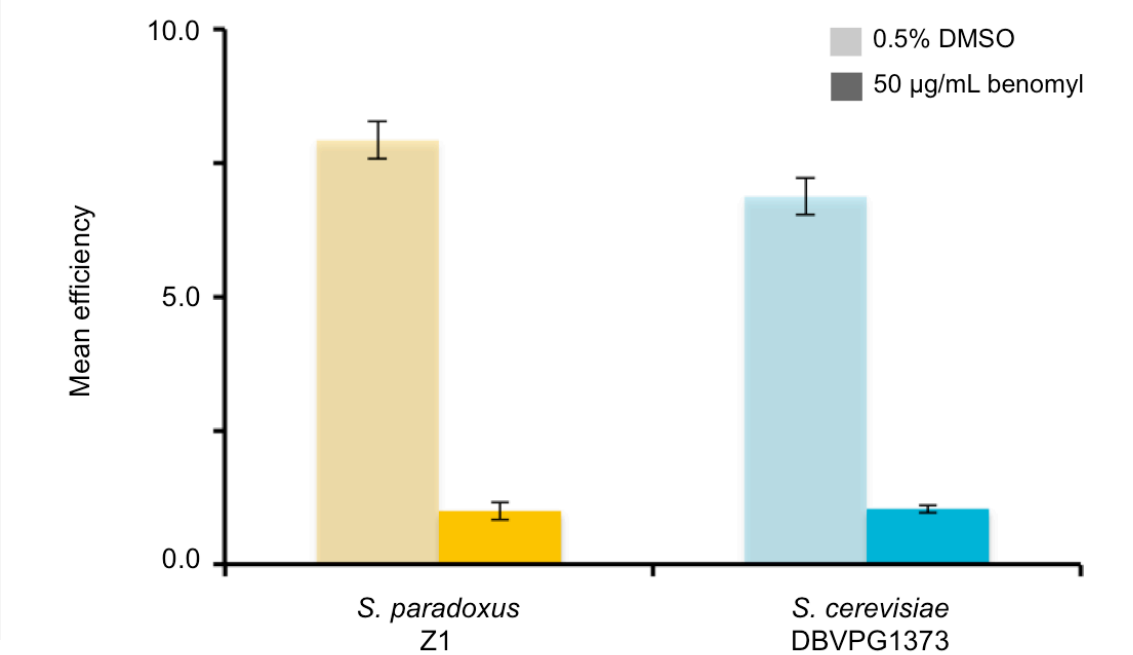


Figure 3. *S. paradoxus* and *S. cerevisiae* have similar resistance to benomyl when grown in liquid culture at 23°C. Each bar represents the mean growth efficiency after 24 hours of the indicated species at 23°C. Light bars show growth in 0.5% DMSO, dark bars show growth in 50 µg/mL benomyl. Means were calculated from n = 2 biological replicate cultures across 1 day's worth of experiments; error bars represent standard deviation.

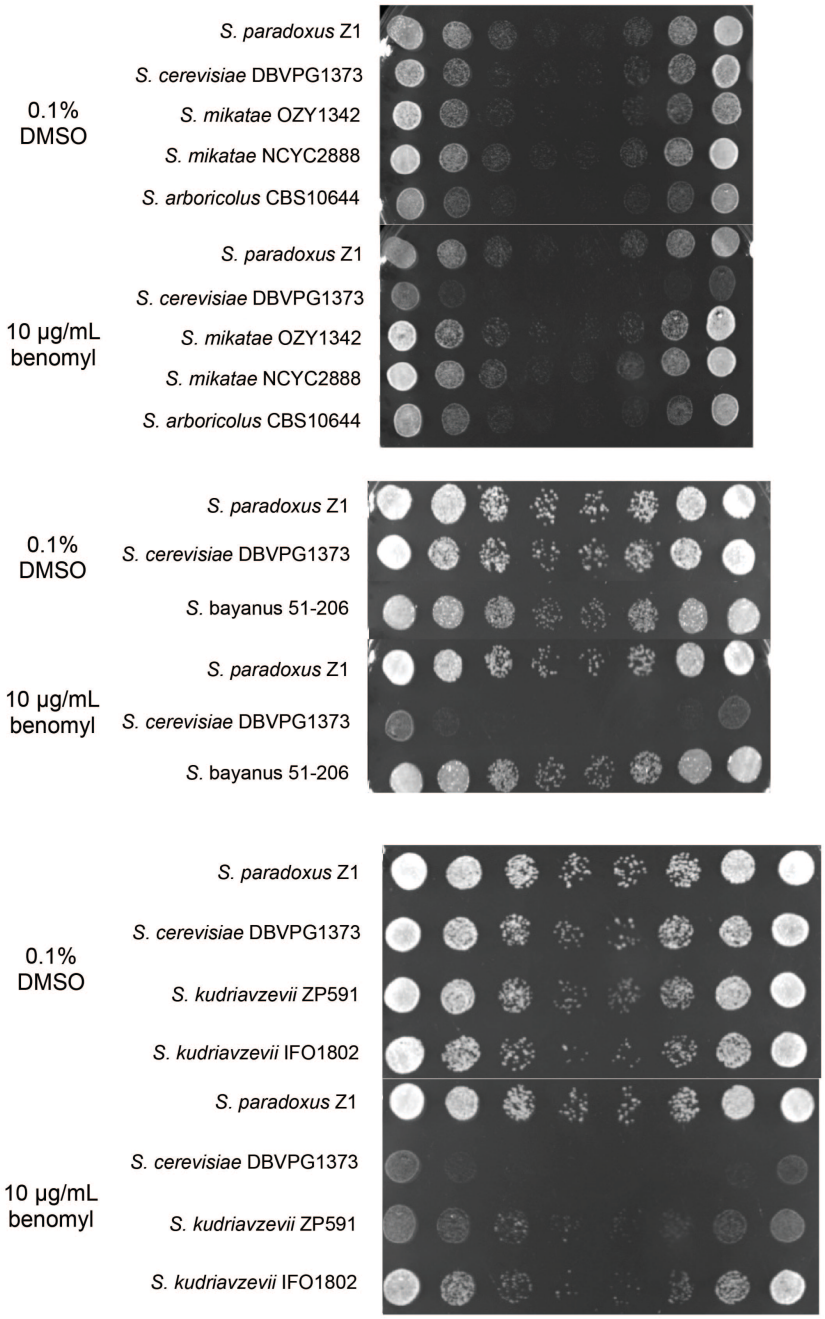
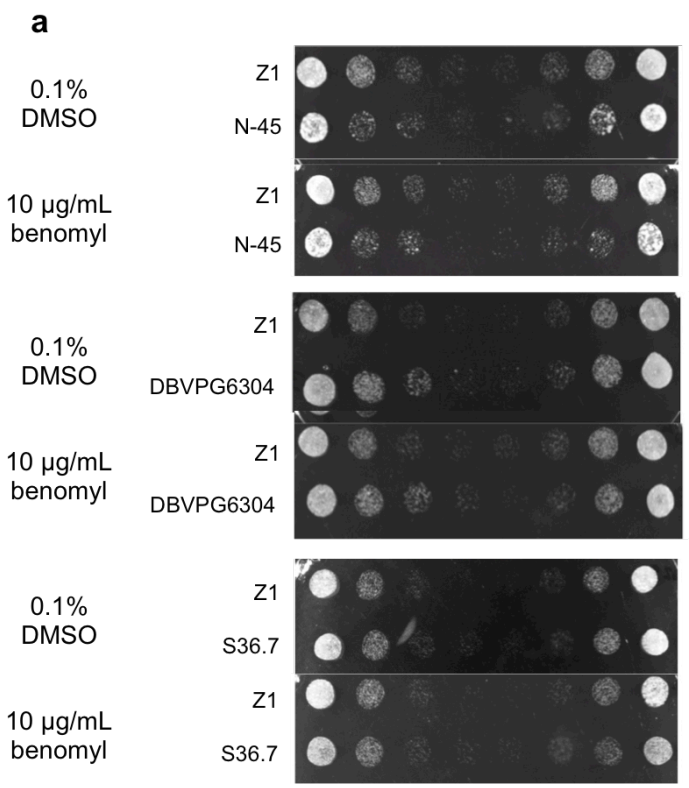


Figure 4. Other *Saccharomyces* species are benomyl resistant. Each image represents a single YPD plate containing either 0.1% DMSO or 10 µg/mL benomyl, as indicated grown for 1-2 days at 23°C. Each dot is liquid culture from the indicated species and strain, starting from concentrated (outermost dots) to less concentrated (innermost dots), moving in 10-fold dilutions. Left and right dilution series represent technical replicates of a dilution series made from the same biological culture.


a
0.1% DMSO
Z1
N-45
10 µg/mL benomyl
Z1
N-45
0.1% DMSO
Z1
DBVPG6304
10 µg/mL benomyl
Z1
DBVPG6304
0.1% DMSO
Z1
S36.7
10 µg/mL benomyl
Z1
S36.7

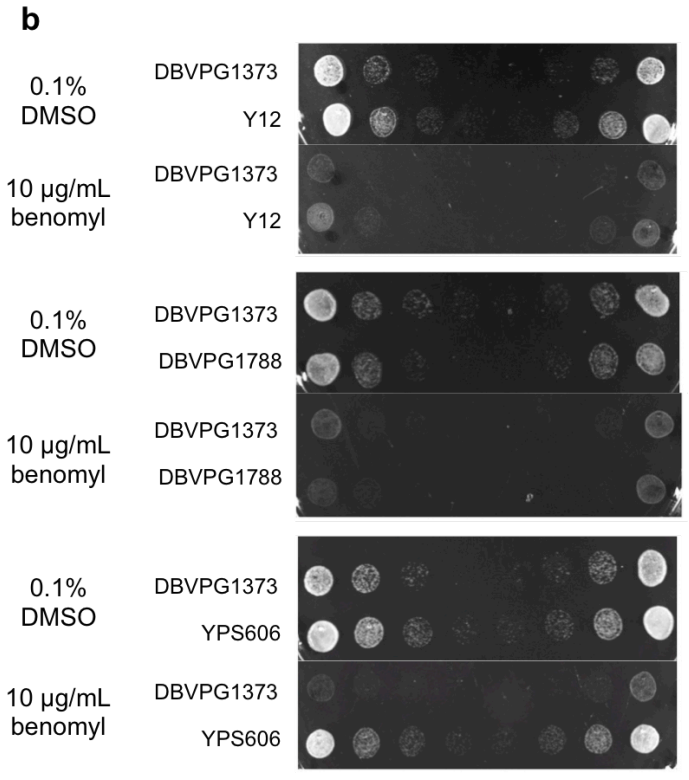

b
0.1% DMSO
DBVPG1373
Y12
10 µg/mL benomyl
DBVPG1373
Y12
0.1% DMSO
DBVPG1373
DBVPG1788
10 µg/mL benomyl
DBVPG1373
DBVPG1788
0.1% DMSO
DBVPG1373
YPS606
10 µg/mL benomyl
DBVPG1373
YPS606

Figure 5. Resistance to benomyl, or lack thereof, is consistent among strains of *S. paradoxus* and *S. cerevisiae.* Each image represents a single YPD plate containing either 0.1% DMSO or 10 µg/mL benomyl, as indicated grown for 1-2 days at 23°C. Each dot is liquid culture from the indicated strain, *S. paradoxus* in **a.** and *S. cerevisiae* in **b.**, respectively, starting from concentrated (outermost dots) to less concentrated (innermost dots), moving in 10-fold dilutions. Left and right dilution series represent technical replicates of a dilution series made from the same biological culture.

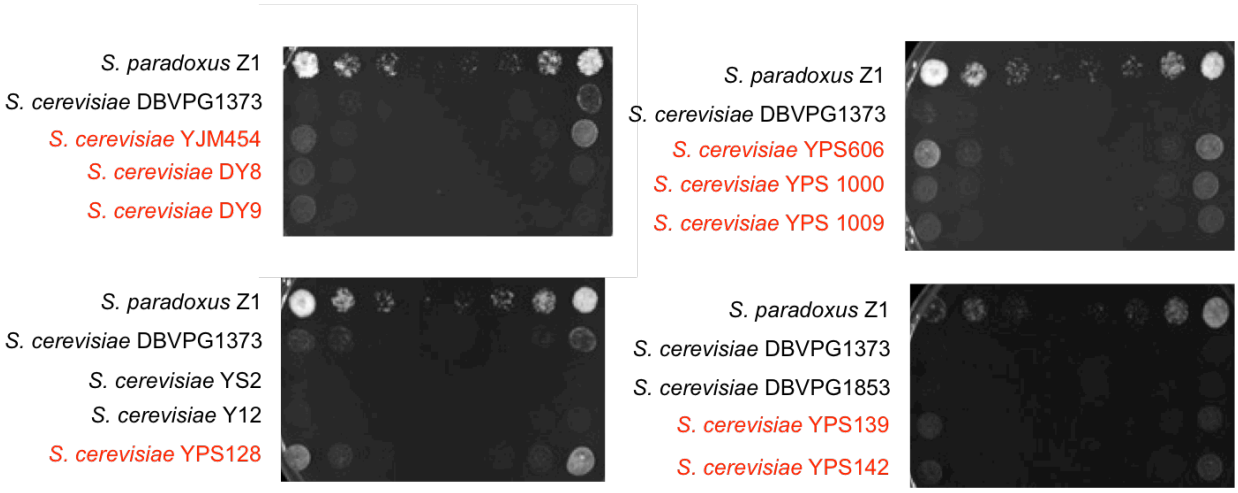
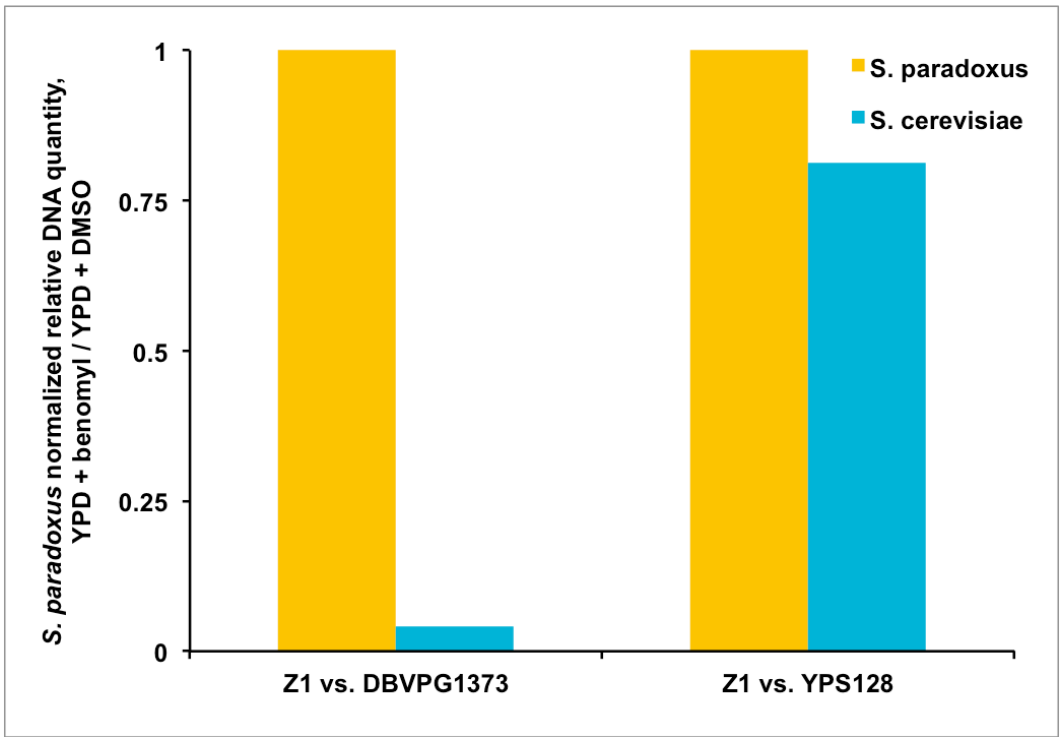


Figure 6. North American *S. cerevisiae* strains have improved cryotolerance relative to other *S. cerevisiae* strains at 4°C. Each image represents a single YPD plate grown for ~1 month at 4°C. Each dot is liquid culture from the indicated species and strain, starting from concentrated (outermost dots) to less concentrated (innermost dots), moving in 10-fold dilutions. Left and right dilution series represent technical replicates of a dilution series made from the same biological culture. North American *S. cerevisiae* strains are in red type. *S. paradoxus* growth is shown for reference.



Supplementary Figure 1. qPCR after competition on solid YPD can quantitatively measure relative benomyl resistance. Each bar represents the DNA quantity of the indicated species (yellow and blue for *S. paradoxus* and *S. cerevisiae*, respectively) and strain in a given competition, after growth on YPD + 10 µg/mL benomyl relative to growth on YPD + 0.1% DMSO, at 23°C, normalized to the quantity of *S. paradoxus* DNA in each competition. Unmatched technical triplicate qPCR reactions were used to calculate each ratio.

Supplementary Tables

Species	Strain name	Details	Source
S. cerevisiae	DBVPG1373	Soil isolate from Netherlands	SGRP
S. paradoxus	Z1	Oak bark isolate from UK	SGRP
S. mikatae	OZY1432	IFO1815 derivative, aka JRY9175	Scannell et. al. [8]
S. mikatae	NCYC2888	aka IFO1815	NCYC
S. arboricolus	CBS10644	Oak bark isolate from China	CBS yeast collection
S. bayanus	51-206	isolated from *Drosphila* from Yosemite, USA	Phaff Yeast Culture Collection
S. kudriavzevii	ZP591	Portugese reference strain	C. Hittinger
S. kudriavzevii	IFO1802	Japanese reference strain	C. Hittinger
S. paradoxus	N-45	Oak tree from Russia	SGRP
S. paradoxus	DBVPG6304	isolated from *Drosophila* from Yosemite	SGRP
S. paradoxus	S36.7	Oak bark isolate from UK	SGRP
S. cerevisiae	Y12	African wine strain	SGRP
S. cerevisiae	DBVPG1788	Soil isolate from Finland	SGRP
S. cerevisiae	YPS606	Oak bark isolate from Pennsylvania, USA	SGRP
S. cerevisiae	YJM454	Clinical isolate from USA	A. Gasch
S. cerevisiae	DY8	Oak tree soil isolate from Wisconsin, USA	A. Gasch
S. cerevisiae	DY9	Oak tree soil isolate from Wisconsin, USA	A. Gasch
S. cerevisiae	YS2	Baker strain from Australia	SGRP
S. cerevisiae	YPS128	Oak soil isolate from Pennsylvania, USA	SGRP
S. cerevisiae	YPS1000	Oak exudate from NJ, USA	A. Gasch
S. cerevisiae	YPS1009	Oak exudate from NJ, USA	A. Gasch
S. cerevisiae	DBVPG1853	Oak soli isolate from Pennsylvania, USA	SGRP
S. cerevisiae	YPS139	Oak soil isolate from Pennsylvania, USA	J. Schacherer
S. cerevisiae	YPS142	Oak soil isolate from Pennsylvania, USA	J. Schacherer

Supplementary Table 1. Strains used in this study. Each row lists the species and strain name of each strain used in this study. The Details column lists information about the geographic location and type of material the strain came from. The Source column notes which culture collection or lab the strain was obtained from.

References

1 Warren, C. D., *et. al.* Distinct chromosome segregation roles for spindle checkpoint proteins. *Mol Biol Cell* **13,** 3029-3041 (2002).

2 Hoyt, M. A., Totis, L. & Roberts, B. T. *S. cerevisiae* genes required for cell cycle arrest in response to loss of microtubule function. *Cell* **66,** 507-517 (1991).

3 Peter, J., *et. al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556,** 339-344 (2018).

4 Davidse, L. C. Benzimidazole fungicides: mechanism of action and biological impact. *Ann Rev Phytopathol* **24,** 43-65 (1986).

5 Wallin, M. & Strömberg, E. Cold-stable and cold-adapted microtubules. *Int Rev Cytol* **157,** 1-31 (1995).

6 Gupta, Jr. M. L., Bode, C. J., Dougherty, C. A., Marquez, R. T. & Himes, R. H. Mutagenesis of β-tubulin cysteine residues in *Saccharomyces cerevisiae:* mutation of cysteine 354 results in cold-stable microtubules. *Cell Motil Cyto* **49,** 67-77 (2001).

7 Will, J. L., Kim, H. S., Clarke, J., Painter, J. C., Fay, J. C. & Gasch, A. P. Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* populations. *PLoS Genetics* **6(4)** e1000893. doi:10.1371/journal.pgen.1000893 (2010).

8. Scannell, D.R. *et. al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the *Saccharomyces sensu stricto* Genus. *G3* **1**, 11-25 (2011).