Title
Unpacking the evolution of thermotolerance in Saccharomyces yeast

Author
Abrams, Melanie Bernette

Publication Date
2022

Peer reviewed|Thesis/dissertation

Unpacking the evolution of thermotolerance in *Saccharomyces* yeast

By

Melanie B. Abrams

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, Chair
Professor John E. Dueber
Professor Kathleen R. Ryan
Professor Benjamin K. Blackman

Spring 2022

Abstract

Unpacking the evolution of thermotolerance in *Saccharomyces* yeast

by

Melanie B. Abrams

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Rachel B. Brem, Chair

A key goal of evolutionary biology is understanding how organisms build new traits. Many adaptations that distinguish taxa evolved so long ago that they are separated across species barriers, and many of these traits also have a complex genetic basis. In this thesis, I have leveraged advances in tools for interspecies genetics to illuminate the genetic underpinnings of thermotolerance evolution in *Saccharomyces* yeast, as a case study. We sought to improve the statistical power of screens to detect genetic determinants of this interspecific trait, both through computational advances in data processing and further development of the screening methodology. We wanted to know what traces of the putative ancient selective sweep(s) on thermotolerance factors might linger in the genomes of modern yeasts. We were also curious how initial conditions the evolutionary landscape might have differed from those further along during the evolutionary acquisition of thermotolerance. In Chapter 1, we introduce the field, and the questions within the field that we studied specifically. In Chapter 2, we investigate the population and comparative genetics of thermotolerance loci, as well as the mechanisms by which one key allele contributes to the phenotype. In Chapter 3, we use a barcoded version of the reciprocal hemizygosity analysis via sequencing screening technique to identify dozens of candidate genes involved in the thermotolerance phenotype. In Chapter 4, we delve into thermotolerance across a range of temperatures and discuss models for the early evolutionary landscape of the thermotolerance divergence. Lastly, in Chapter 5, we use directed evolution to gain insight into adaptations that allow a thermosensitive *Saccharomyces* yeast to acquire thermotolerance.

Chapter 1

An introduction to the study of the evolution of complex, ancient traits, and *Saccharomyces* thermotolerance as a model system for this field

The study of the evolution of complex, ancient traits

Life on earth emerged billions of years ago (Pearce et al., 2018); all the organisms we encounter today have been shaped by evolution over the eons since. Now, we live surrounded by organisms that are marvelously adapted to their specific niches—from mosses that perpetuate using the small gaps on logs left by chipmunks (Kimmerer and Young, 1996) to arctic woolly caterpillars which survive repeatedly freezing and thawing before they pupate and turn into moths (Kukal, 1991). Organisms' genomes contain the DNA encoding their diverse phenotypes, as well as the molecular signals of past and present pressure from natural selection and traces of genetic drift. As evolutionary biologists, we want to understand the processes that have shaped the natural world, and we want to understand the genetic underpinnings of the changes that evolve in the wild.

Over time, populations of related organisms diverge so much that they are then considered separate species. In sexually reproducing organisms, speciation corresponds with reproductive isolation. It becomes highly unlikely, and often impossible, for members of the two new species to interbreed. Species can be pre-zygotically isolated, meaning that they do not form zygotes at all (through mechanisms such as behavioral isolation, like different mating songs in birds, or gametic isolation, where the gametes do not successfully meet and merge to form hybrids)(Ludlow and Magurran, 2006; Marshall and Dirienzo, 2012; Uy et al., 2018). Species may also be post-zygotically isolated, meaning that they either hybrids less fit than both parents, fully inviable hybrids, or meiotically infertile hybrids, like the mule offspring of donkeys and horses, which can live but not reproduce (Chandley et al., 1974; Islam et al., 2013; Rometsch et al., 2020). With complete reproductive isolation, each of the new species will continue to evolve, but independently—and new traits that arise after speciation are separated by the species barrier.

Any time a trait with a complex genetic basis manifests between isolated species, it is a challenge to understand its genetic basis. There are a variety of tools that have helped scientists develop an understanding of the molecular mechanisms which govern local adaptation among populations of the same species in eukaryotes (Nadeau and Jiggins, 2010; Savolainen et al., 2013), but those classical statistical genetics screens rely on meiotic recombination in the hybrid for linkage analysis. Those tools, like GWAS and QTL mapping, cannot be applied to traits separated by species barriers, because even viable interspecific hybrids do not successfully undergo meiosis.

Candidate gene and comparative genomic approaches have proven invaluable to the field of interspecific genetics, largely by examining cases of a single large-effect size determinant of a trait. Candidate gene approaches begin with an educated guess based

on the relevance of a gene's known functions to the phenotypic variation of interest, and then follow up with experimental validation of that guess (Spiewak et al., 2018; Sulak et al., 2016; Xie et al., 2018). For example, ancestral protein reconstruction on candidate genes can illuminate the molecular mechanisms of an ancient divergence in the function of a particular protein (Anderson et al., 2015; Bridgham et al., 2009). These studies advance our understanding of how traits evolved deep in that past, but these methods typically characterize a single molecular determinant for the phenotype in question. As a result, there remains a gap in knowledge in the field of interspecific genetics surrounding the molecular mechanisms which govern polygenic adaptation over long time periods.

Experimental evolution and theoretical predictions offer insight into potential mechanisms which may play roles in complex, ancient, adaptive evolutionary processes. Under a theoretical framework proceeding from Fisher's geometric model of adaptation (Fisher, 1930), we expect most traits are polygenic (Orr, 1998). Laboratory evolution of *Escherichia coli* suggests that historical contingency may be an important factor in long-term adaptation (Blount et al., 2018). These paradigms have led to compelling models for how evolution works in the natural world, and they create an opening to further the field of evolutionary biology through empirical validation of those models in systems which have undergone adaptive processes in the wild.

In the past few years, new techniques have been developed that offer the opportunity to map the genetic basis of polygenic, interspecific phenotypes. These new methods rely on screening interspecific hybrids (Lazzarano et al., 2018; Sadhu et al., 2016; Sharon et al., 2018; Weiss et al., 2018a), which is possible using cell culture that involves mitotic, but not meiotic, propagation of the hybrid line. Unbiased genetic screens for factors underlying ancient traits can reveal their complex genetic underpinnings, and further study of those factors can shed light on the mechanisms involved in the evolutionary processes.

Saccharomyces thermotolerance as a model for studying complex, ancient adaptation

The thermotolerance trait in *Saccharomyces* yeast offers a compelling model system for studying how evolution builds traits over long periods of time. *Saccharomyces cerevisiae*, also known as baker's yeast or brewer's yeast, is a well-annotated, genetically tractable, fast-growing model organism. *S. cerevisiae* diverged from its sister species, *Saccharomyces paradoxus*, ~5 million years ago (Shen et al., 2018). *S. cerevisiae* and *S. paradoxus* form mitotically viable hybrids, allowing screens in the hybrid background.

S. cerevisiae is the most thermotolerant member of its clade, a putatively adaptive trait (Gonçalves et al., 2011; Salvadó et al., 2011; Sweeney et al., 2004; Weiss et al., 2018a). *S. cerevisiae* can grow and accumulate biomass at 39°C, while its sister species *S. paradoxus* exhibits the putatively ancestral phenotype and fails to accumulate much biomass at that temperature (Weiss et al., 2018a). Some strains of *S.*

cerevisiae have been documented to grow at temperatures as high as 41°C (Gonçalves et al., 2011); conversely, strains of *S. paradoxus* show a drastic loss of viability at temperatures in the range of 36-37°C (AlZaben et al., 2021). *S. cerevisiae* is believed to have diverged from other *Saccharomyces* in East Asia (Peter et al., 2018) and, potentially, could have acquired thermotolerance under selective pressures exerted by such a hot niche. Modern *S. cerevisiae* and *S. paradoxus* often inhabit similar habitats in the wild, but sympatric populations of the two species still have markedly different thermotolerance phenotypes (Sweeney et al., 2004). Under one compelling model, *S. cerevisiae* may have acquired thermotolerance and ethanol tolerance in tandem, to gain a competitive advantage in fermenting fruit where the yeast's rapid growth produced heat and alcohol (Goddard, 2008; Salvadó et al., 2011).

Our laboratory previously used the thermotolerance divergence between *S. cerevisiae* and *S. paradoxus* as a case for interspecific genetic study using a technique called reciprocal hemizygosity analysis via sequencing (RH-seq) (Weiss et al., 2018a). The genes uncovered by this unbiased genetic screen were largely housekeeping genes, many of which are essential factors in chromosome segregation and mitosis. No single *S. cerevisiae* allele was sufficient to recapitulate the full thermotolerance phenotype in an otherwise thermosensitive *S. paradoxus* background, pointing to the complexity of the trait. Somehow, in the millions of years since *S. cerevisiae* and *S. paradoxus* diverged, the ancestor of modern *S. cerevisiae* had made changes to a variety of essential genes, while preserving those genes' essential functions. That struck us as a massive evolutionary challenge, particularly given how much harder it is to acquire multiple unlinked variants in the same background to achieve a phenotype than it is to change a single allele (Lynch, 2010; Lynch and Abegg, 2010).

The partially characterized genetics of the thermotolerance divergence between *S. cerevisiae* and other *Saccharomyces* offered fertile territory for a deeper investigation of this case of an ancient, complex adaptation. We wanted to know more about the scale and scope of the evolutionary challenge that the ancestor of modern *S. cerevisiae* faced in acquiring thermotolerance.

The contents of this thesis

The following chapters describe insights into a variety of facets of the evolution of thermotolerance in *S. cerevisiae*.

We wanted to know whether the evolution of thermotolerance factors was driven by positive selection between species, and what had happened to the thermotolerant alleles as modern *S. cerevisiae* populations diverged. We also wanted to know more about the mechanism by which a *S. cerevisiae* allele might contribute to the thermotolerance phenotype. In the next chapter, we used population and comparative-genetic tools to learn about the evolutionary history of thermotolerance loci from the sequences of modern *S. cerevisiae* and *S. paradoxus*. We started by re-analyzing the raw data from the RH-seq screen of the impact of variation between the species on thermotolerance from (Weiss et al., 2018a) with a modified analysis pipeline that

improves on the statistical power to detect thermotolerance factors. We used the resulting hits as input into an analysis of metrics for divergence, amino acid changes, and conservation, which paint a picture of divergence between species, and conservation within the *S. cerevisiae* branch, at the thermotolerance loci. Next, we took *ESP1*, a particularly strong thermotolerance determinant, as a case for closer inspection. We used microscopy to validate a hypothesis that, at high temperature, yeast cells harboring the thermosensitive *S. paradoxus* allele of *ESP1* die in the process of cell division—a phenomenon consistent with the prevalence of mitosis genes among the identified thermotolerance loci.

After that, we wanted to get a sense of how many thermotolerance factors we could detect—whether we already had in hand the largest determinants of the phenotype, or if we could find a wider set. Then, we wanted to know whether that wider set would also show a pattern of being involved in chromosome segregation and mitosis—or if other cellular processes were important in the *Saccharomyces* thermotolerance divergence. In the next chapter, we used barcoded transposons in an RH-seq screen, which makes it cheaper and easier to screen for genetic determinants of an interspecific phenotype. The new method allowed us to perform the screen with many more replicates, leading to higher statistical power, which in turn permitted us to detect an order of magnitude more candidate thermotolerance loci than the lab's earlier non-barcoded approach (Weiss et al., 2018a). We found functional enrichment for mitosis genes, and other housekeeping factors from different cellular pathways. We also found loci with robust signals of positive selection among the candidate loci.

These results suggested a vastly complex, if not omnigenic, genetic architecture for the thermotolerance trait. We were curious about how this complex set of genetic determinants might allow inferences of the evolutionary landscape of the ancestor of modern *S. cerevisiae* as it began to acquire thermotolerance. In particular, we were interested in investigating evidence for a model of adaptation by the *S. cerevisiae* lineage along a temperature cline, with ancestral populations gradually expanding their range from temperate to warmer and warmer conditions. To pursue this notion, we used *S. paradoxus*, the sister species of *S. cerevisiae,* as a proxy for the ancestor of modern *S. cerevisiae* with respect to thermotolerance. Using that proxy, viability data suggested that the ancestor of *S. cerevisiae* likely would have needed to begin adaptation towards thermotolerance at a temperature well below where it can now thrive. We then examined the temperature-responsive growth of *S. cerevisiae* strains harboring alleles from *S. paradoxus* at thermotolerance loci. Some of the latter broke down at lower temperatures, where we believe thermotolerance adaptation likely initiated in the wild ancestor of modern *S. cerevisiae*. Some only manifested phenotypically at higher temperatures. Our working model is thus that ancestral alleles at thermotolerance loci would have been subject to selective pressure for thermotolerance adaptation at different stages of adaptation across the cline from temperate to warmer temperatures.

We were curious about the initial stages of an evolutionary process whereby a thermosensitive *Saccharomyces* might acquire thermotolerance, particularly because initial genomic changes in *S. cerevisiae* may have been transient, and thus untraceable

in modern strains of the species. Again using *S. paradoxus* as a proxy for the ancestor of *S. cerevisiae* again, we performed directed evolution with a consistent selective pressure to improve thermotolerance. This program led to thermotolerant strains of *S. paradoxus*. Investigating what might have caused this heritable thermotolerance improvement, we found evidence of several chromosomal duplications in the evolved strains.

Together, this work adds detail and depth to our understanding of how modern *S. cerevisiae* could have acquired its exceptional thermotolerance. We hope it will provide a useful case study in a model system for the growing field of study of complex, ancient adaptations in the wild.

REFERENCES

AlZaben, F., Chuong, J.N., Abrams, M.B., and Brem, R.B. (2021). Joint effects of genes underlying a temperature specialization tradeoff in yeast. PLoS Genet 17, e1009793.

Anderson, D.W., McKeown, A.N., and Thornton, J.W. (2015). Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. ELife 4, e07864.

Blount, Z.D., Lenski, R.E., and Losos, J.B. (2018). Contingency and determinism in evolution: Replaying life's tape. Science 362, eaam5979.

Bridgham, J.T., Ortlund, E.A., and Thornton, J.W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461, 515–519.

Chandley, A.C., Jones, R.C., Dott, H.M., Allen, W.R., and Short, R.V. (1974). Meiosis in interspecific equine hybrids. CGR 13, 330–341.

Fisher, R.A. (1930). The genetical theory of natural selection (Рипол Классик).

Goddard, M.R. (2008). Quantifying the complexities of Saccharomyces cerevisiae's ecosystem engineering via fermentation. Ecology 89, 2077–2082.

Gonçalves, P., Valério, E., Correia, C., Almeida, J.M.G.C.F. de, and Sampaio, J.P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric Saccharomyces Species. PLOS ONE 6, e20739.

Islam, F.B., Ishishita, S., Uno, Y., Mollah, M.B.R., Srikulnath, K., and Matsuda, Y. (2013). Male Hybrid Sterility in the Mule Duck is Associated with Meiotic Arrest in Primary Spermatocytes. The Journal of Poultry Science advpub.

Kimmerer, R.W., and Young, C.C. (1996). Effect of Gap Size and Regeneration Niche on Species Coexistence in Bryophyte Communities. Bulletin of the Torrey Botanical Club 123, 16–24.

Kukal, O. (1991). Behavioral and Physiological Adaptations to Cold in a Freeze-Tolerant Arctic Insect. In Insects at Low Temperature, R.E. Lee, and D.L. Denlinger, eds. (Boston, MA: Springer US), pp. 276–300.

Lazzarano, S., Kučka, M., Castro, J.P.L., Naumann, R., Medina, P., Fletcher, M.N.C., Wombacher, R., Gribnau, J., Hochepied, T., Montagu, M.V., et al. (2018). Genetic mapping of species differences via in vitro crosses in mouse embryonic stem cells. PNAS 115, 3680–3685.

Ludlow, A. m, and Magurran, A. e (2006). Gametic isolation in guppies (Poecilia reticulata). Proceedings of the Royal Society B: Biological Sciences 273, 2477–2482.
Lynch, M. (2010). Scaling expectations for the time to establishment of complex adaptations. Proc Natl Acad Sci U S A 107, 16577–16582.
Lynch, M., and Abegg, A. (2010). The Rate of Establishment of Complex Adaptations. Mol Biol Evol 27, 1404–1414.
Marshall, J.L., and Dirienzo, N. (2012). Noncompetitive Gametic Isolation between Sibling Species of Cricket: A Hypothesized Link between Within-Population Incompatibility and Reproductive Isolation between Species. Int J Evol Biol 2012, 593438.
Nadeau, N.J., and Jiggins, C.D. (2010). A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Genet. 26, 484–492.
Orr, H.A. (1998). The Population Genetics of Adaptation: The Distribution of Factors Fixed During Adaptive Evolution. Evolution 52, 935–949.
Pearce, B.K.D., Tupper, A.S., Pudritz, R.E., and Higgs, P.G. (2018). Constraining the Time Interval for the Origin of Life on Earth. Astrobiology 18, 343–364.
Peter, J., Chiara, M.D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344.
Rometsch, S.J., Torres-Dowdall, J., and Meyer, A. (2020). Evolutionary dynamics of pre- and postzygotic reproductive isolation in cichlid fishes. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190535.
Sadhu, M.J., Bloom, J.S., Day, L., and Kruglyak, L. (2016). CRISPR-directed mitotic recombination enables genetic mapping without crosses. Science 352, 1113–1116.
Salvadó, Z., Arroyo-López, F.N., Guillamón, J.M., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus Saccharomyces. Appl Environ Microbiol 77, 2292–2302.
Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics 14, 807–820.
Sharon, E., Chen, S.-A.A., Khosla, N.M., Smith, J.D., Pritchard, J.K., and Fraser, H.B. (2018). Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. Cell 175, 544-557.e16.
Shen, X.-X., Opulente, D.A., Kominek, J., Zhou, X., Steenwyk, J.L., Buh, K.V., Haase, M.A.B., Wisecaver, J.H., Wang, M., Doering, D.T., et al. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell 175, 1533-1545.e20.
Spiewak, J.E., Bain, E.J., Liu, J., Kou, K., Sturiale, S.L., Patterson, L.B., Diba, P., Eisen, J.S., Braasch, I., Ganz, J., et al. (2018). Evolution of Endothelin signaling and diversification of adult pigment pattern in Danio fishes. PLOS Genetics 14, e1007538.
Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D., and Lynch, V.J. (2016). TP53 copy number expansion is associated with the

evolution of increased body size and an enhanced DNA damage response in elephants. ELife 5, e11994.
Sweeney, J.Y., Kuehne, H.A., and Sniegowski, P.D. (2004). Sympatric natural Saccharomyces cerevisiae and S. paradoxus populations have different thermal growth profiles. FEMS Yeast Res. 4, 521–525.
Uy, J.A.C., Irwin, D.E., and Webster, M.S. (2018). Behavioral Isolation and Incipient Speciation in Birds. Annual Review of Ecology, Evolution, and Systematics 49, 1–24.
Weiss, C.V., Roop, J.I., Hackley, R.K., Chuong, J.N., Grigoriev, I.V., Arkin, A.P., Skerker, J.M., and Brem, R.B. (2018). Genetic dissection of interspecific differences in yeast thermotolerance. Nature Genetics 50, 1501.
Xie, J., Li, Y., Shen, X., Goh, G., Zhu, Y., Cui, J., Wang, L.-F., Shi, Z.-L., and Zhou, P. (2018). Dampened STING-Dependent Interferon Activation in Bats. Cell Host & Microbe 23, 297-301.e4.

Chapter 2

Population and comparative genetics of thermotolerance divergence between yeast species.

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

Abrams, M.B., Dubin, C.A., AlZaben, F., Bravo, J., Joubert, P.M., Weiss, C.V., and Brem, R.B. (2021). Population and comparative genetics of thermotolerance divergence between yeast species. G3 Genes|Genomes|Genetics.

ABSTRACT

Many familiar traits in the natural world—from lions' manes to the longevity of bristlecone pine trees—arose in the distant past, and have long since fixed in their respective species. A key challenge in evolutionary genetics is to figure out how and why species-defining traits have come to be. We used the thermotolerance growth advantage of the yeast *Saccharomyces cerevisiae* over its sister species *Saccharomyces paradoxus* as a model for addressing these questions. Analyzing loci at which the *S. cerevisiae* allele promotes thermotolerance, we detected robust evidence for positive selection, including amino acid divergence between the species and conservation within *S. cerevisiae* populations. Since such signatures were particularly strong at the chromosome segregation gene *ESP1*, we used this locus as a case study for focused mechanistic follow-up. Experiments revealed that, in culture at high temperature, the *S. paradoxus ESP1* allele conferred a qualitative defect in biomass accumulation and cell division relative to the *S. cerevisiae* allele. Only genetic divergence in the *ESP1* coding region mattered phenotypically, with no functional impact detectable from the promoter. Together, these data support a model in which an ancient ancestor of *S. cerevisiae,* under selection to boost viability at high temperature, acquired amino acid variants at *ESP1* and many other loci, which have been constrained since then. Complex adaptations of this type hold promise as a paradigm for interspecies genetics, especially in deeply diverged traits that may have taken millions of years to evolve.

INTRODUCTION

A central goal of research in evolutionary genetics is to understand how new traits are built. Much of the literature to date focuses on adaptive trait innovation within a species, in the wild (Asgari et al., 2020; Chan et al., 2010; Cleves et al., 2014; Field et al., 2016; Linnen et al., 2013; Will et al., 2010) and in the lab (Blount et al., 2012; Castro et al., 2019; Good et al., 2017; Tenaillon et al., 2016). These systems have enabled studies of short-term adaptation, its genetics (Barroso-Batista et al., 2014; Castro et al., 2019; Garud et al., 2015; Good et al., 2017; Harris et al., 2018; Xie et al., 2019) and its dynamics (Blount et al., 2012; Toprak et al., 2011). Such work on recent adaptations serves as a backdrop for the study of evolution over longer timescales. Many familiar

traits from the natural world have been acquired over millions of generations. In the modern day, such characters manifest as differences between deeply diverged, reproductively isolated lineages. They can represent the abiding fitness strategies of their respective species, and are thus of particular interest in the field. But their evolutionary mechanisms pose a key challenge, given that the relevant events happened so long ago. For these ancient traits, candidate-gene studies have implicated individual loci (Anderson et al., 2016; Baldwin et al., 2014; Li and Fay, 2017; Liu et al., 2018; Massey and Wittkopp, 2016; Sackton et al., 2019; Sulak et al., 2016; Tian et al., 2019) and reconstructed the mutational path by which a given determinant evolved (Anderson et al., 2015; Bridgham et al., 2009; Finnigan et al., 2012; Liu et al., 2018; Pillai et al., 2020). Even in such landmark cases, the tempo and mode of evolution of deep trait divergences have remained largely out of reach. To meet the latter challenge, one would need to trace the rise of causal alleles in the respective species and the selective forces that drove it, and pinpoint the timing of these events.

In prior work, our group mapped multiple housekeeping genes underlying the difference in thermotolerance between *Saccharomyces cerevisiae* and other species in its clade, and found that *S. cerevisiae* harbored derived alleles at these loci (Weiss et al., 2018a). Here we set out to investigate when and how *S. cerevisiae* acquired the putatively adaptive determinants of thermotolerance, using a population-genomic approach. We then used the results as a jumping-off point for additional analyses of the molecular mechanisms by which variants at thermotolerance genes confer their effects.

MATERIALS AND METHODS

Identifying thermotolerance genes

To identify genes at which variation between *S. cerevisiae* and *S. paradoxus* impacts thermotolerance, we re-analyzed data from a reciprocal hemizygosity screen of transposon mutants in the interspecies hybrid background (Weiss et al., 2018a) as described, with the following differences. Call $a_{39,i}$ the average, across technical replicates, of sequencing-based abundances of a hemizygote mutant measured after ~7 generations in biological replicate *i* of growth at 39°C, and $a_{28,i}$ the analogous quantity for growth at 28°C, for i = [1,3]. We calculated the mean of the latter across biological replicates, $a_{28,mean}$, and then used it to tabulate three replicate estimates of the temperature effect on growth of the mutant as $\log_2(a_{39,i}/a_{28,mean})$. If the coefficient of variation across these biological replicates was greater than 20, we eliminated the mutant from further consideration. Otherwise, for a given gene, we concatenated these vectors of length three across all hemizygote mutants in the *S. paradoxus* allele for which we had abundance data, yielding the set of temperature effects s_{Spar}. We did likewise for the *S. cerevisiae* allele, yielding s_{Scer}. We retained for further analysis only genes at which we had at least two mutants' worth of data for each allele. For each such gene, we compared s_{Scer} and s_{Spar} with a Wilcoxon test, and corrected for multiple testing across genes, as described in (Weiss et al., 2018a).

Sequence data, alignments, and interspecies diversity

For D_{XY} analyses in Table S2, for a given gene, open reading frame sequences for the strains of each *S. cerevisiae* population from (Peter et al., 2018) were aligned against the European *S. paradoxus* population from (Bergström et al., 2014) and, separately, against the North American *S. paradoxus* subpopulation B from (Durand et al., 2019). For D_{XY} analysis across species in Table 1, alignments were generated using all *S. cerevisiae* and *S. paradoxus* strains. Alignments used MUSCLE (Edgar, 2004) with the default settings for DNA and --maxiters set to 2. Any gene for which, in the alignment, >10% of sites were denoted as gaps or unknown nucleotides (Ns), or sequences from <75% of strains in the population were available, was eliminated from analysis, leaving 4110 to 4781 genes suitable for testing in each analysis.

We calculated pairwise nucleotide diversity (D_{xy}) for each gene as
$D_{XY} = \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} d_{ij}$
using a custom script, where n_x is the number of *S. cerevisia*e strains, n_y is the number of *S. paradoxus* strains, and *d* is the number of sites with nucleotides differences at the same position for each pairwise sequence comparison. Sites with gaps or unknown nucleotides were ignored.

To test for enriched D_{xy} among thermotolerance genes in a given *S. cerevisiae* and *S paradoxus* population pair or across the species, we first tabulated D_{therm}, the median D_{xy} across the thermotolerance gene cohort from the appropriate genomes. We next sampled 10,000 random cohorts of genes from the genome with the same number of essential and nonessential genes as in the thermotolerance cohort (Winzeler et al., 1999), and tabulated the median D_{xy} in each D_{rand} from the appropriate genomes. We used as an empirical *p*-value the proportion of random cohorts with $D_{rand} \geq D_{therm}$.

Codon alignment and McDonald-Kreitman statistics

Open reading frame sequences for each *S. cerevisiae* strain from (Peter et al., 2018), the European *S. paradoxus* strains (Bergström et al., 2014), and North American *S. paradoxus* strains (Durand et al., 2019) were translated to amino acid sequences using Biopython (Cock et al., 2009) and aligned using MUSCLE (Edgar, 2004) with default settings for amino acids and --maxiters set to 2. The amino acid sequence alignments and unaligned nucleotide sequences were used as input to PAL2NAL (Suyama et al., 2006) to create codon alignments for each gene. Sequences with stop codons within the open reading frame or where >10% of sites were denoted as gaps or unknown nucleotides (Ns) were discarded. Genes with valid sequences in <75% of strains from each species were removed from the analysis, leaving 3814 genes suitable for testing.

The codon alignments were input into the CodonAlignment module of Biopython 1.78 (Cock et al., 2009) and the mktest function reported the number of divergent nonsynonymous (D_n), divergent synonymous (D_s), polymorphic nonsynonymous (P_n), and polymorphic synonymous changes (P_s) in each gene. We calculated the Neutrality Index (NI) for each gene as

$$\mathrm{NI} = \frac{P_n / P_s}{D_{n/} D_s}$$

(McDonald and Kreitman, 1991). We then used these measures as input into a resampling test for enrichment of low NI, analogous to that used for D_{xy} (see above).

Multi-locus genotype and allele-sharing inference in *S. cerevisiae* populations

We calculated expected genotype homozygosity, G1 (Harris et al., 2018), as follows. For the allele-sharing inference across all *S. cerevisiae* in Table 1, we used unphased VCF genotypes for all strains from 1011 Yeast Genomes (Peter et al., 2018) as input into SelectionHapStats (Harris et al., 2018) with the following parameters: -w (window size, SNPs) = 1200, and -j (jump size, SNPs) = 25. We tabulated G1 in each window whose center fell within the open reading frame, and we calculated the average across the windows. We then used these measures as input into a resampling test for enrichment of high G1 analogous to that used for D_{xy} (see above). For allele-sharing inference in individual populations of *S. cerevisiae* in Table S3, we proceeded as above except that we used unphased VCF genotypes for each of the five largest *S. cerevisiae* populations from 1011 Yeast Genomes (Peter et al., 2018).

We also evaluated allele-sharing at thermotolerance genes in *S. paradoxus*, as a complement to the above analyses in *S. cerevisiae*. For this purpose we used unphased VCF genotypes for all *S. paradoxus* genomes from (Bergström et al., 2014) according to the methods above and found no significant enrichment for G1 at thermotolerance loci (resampling $p = 0.19$).

***ESP1* phylogenetic analysis**

We used the alignment of the open reading frame of *ESP1* from the type strains of *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. bayanus*, *S. uvarum*, and *S. kudriavzevii* from saccharomycessensustricto.org as input into the codeml module of PAML4.9 (Yang, 2007). The branch-site model (model=2, NSsites=2) was used, and two models, null and alternative, were fitted. In the null model, the dN/dS for the *S. cerevisiae* branch was fixed at 1.0 and all other branches were described by the same dN/dS ratio (ω). In the alternative model, the branch leading to *S. cerevisiae* was fitted with one ω, and all other branches were fitted with a separate ω. A test statistic, calculated by comparing the likelihood ratios of the alternative and null models, was used to calculate a *p*-value by comparing it to a chi-squared distribution with one degree of freedom, equal to the difference in the number of parameters in the two models. No codons exhibited a posterior probability of positive selection, on the branch leading to *S. cerevisiae*, higher than 0.9.

Analysis of *cis*-regulatory expression divergence between *S. cerevisiae* and *S. paradoxus*

For Table S5, we analyzed temperature-dependent allele-specific expression measurements in interspecific hybrids as follows. For each gene in turn, from (Tirosh et

al., 2009) we tabulated the $\log_2$-ratio of allele-specific expression between alleles of an *S. cerevisiae* x *S. paradoxus* hybrid cultured at 35°C, as a difference from the analogous quantity from cultures at 30°C, which we refer to as D_{ASE}. To test for an enrichment of high-magnitude allele-specific expression differences between species at thermotolerance loci, we took the absolute value of D_{ASE} for each gene and then tabulated the median of this value across the thermotolerance gene set, $D_{ASE,therm}$. We next sampled 10,000 random cohorts of genes from the genome with the same number of essential and nonessential genes as in the thermotolerance cohort (Winzeler et al., 1999), and for each we calculated the median D_{ASE}, $D_{ASE,rand}$. We then used as an empirical *p*-value the proportion of random cohorts for which $D_{ASE,rand} \geq D_{ASE,therm}$. To test for directional *cis*-regulatory change between species at thermotolerance genes, we repeated the above analysis except that we took the median across signed D_{ASE} values for a gene set of interest, and then used as an empirical *p*-value the proportion of random cohorts at which $|ASE_{rand}| \geq |ASE_{therm}|$.

Separately, we repeated the above analysis using measurements from 37°C and 33°C cultures of an *S. cerevisiae* x *S. uvarum* hybrid from (Li and Fay, 2017), which are reported as allele-specific expression $a_{T,s}$ for the allele from species *s* at temperature *T*. We tabulated D_{ASE} for each gene as $\log_2(a_{37,Scer}/a_{37,Suv}) - \log_2(a_{33,Scer}/a_{33,Suv})$ and tested for enrichment of high-magnitude and directional *cis*-regulatory variation across thermotolerance genes as above.

Interspecies swap strain construction at *ESP1* promoter and coding region

To swap the allele of the *ESP1* promoter from *S. paradoxus* Z1 into *S. cerevisiae* DBVPG1373, and likewise for the coding region, we designed allele-specific Cas9 guide RNAs for the *S. cerevisiae* background, generated donor DNA from *S. paradoxus*, transformed, and screened for successful transgenesis by Sanger sequencing as in (Weiss et al., 2018a). Strains are listed in Table S6.

Large-format growth assay

For growth measurements in Figures 2 and S2, we assayed *S. paradoxus* Z1, *S. cerevisiae* DBVPG1373, the full *ESP1* swap in the *S. cerevisiae* background (harboring the promoter and open reading frame from *S. paradoxus*) from (Weiss et al., 2018a), and the *ESP1* promoter and coding swaps in the *S. cerevisiae* background (see above) as follows. Each strain was streaked from a -80°C freezer stock onto a yeast peptone dextrose (YPD) agar plate and incubated at room temperature for 3 days. For each biological replicate, a single colony was inoculated into 5 mL liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate pre-cultures. Each pre-culture 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. Each pre-culture was again back-diluted into 10 mL 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 either 28 or 39°C with shaking at 200rpm for 24 hours, and OD_{600} was measured again. The growth efficiency for each replicate was calculated as the difference between these final and

initial OD_{600} values. The pipeline from inoculation off of solid plates through preculture, two back-dilutions, and growth at 28 or 39°C we refer to as a day's growth experiment. For each day's experiments, we calculated the average efficiency $\langle e_{Scer} \rangle$ across the replicates of wild-type *S. cerevisiae*, and we used this quantity to normalize the efficiency e_s measured for each replicate assayed on that day of a genotype of interest *s*. Thus, the final measurement used for analysis for each replicate on a given day was $e_s/\langle e_{Scer} \rangle$. We carried out a total of 2-3 days' worth of replicate growth experiments for each genotype, with three separate transformant strains analyzed by this workflow in the case of the coding swap. For a given genotype we used the complete cohort of measurements of $e_s /\langle e_{Scer} \rangle$ from all days and strains as input into a one-sample, one-tailed Wilcoxon test to evaluate whether $e_s/\langle e_{Scer} \rangle$ was less than 1 (*i.e.* that the strain grew worse at 39°C than wild-type *S. cerevisiae*).

Temperature dose-response growth assay

To evaluate temperature dose-responses in Figure 3, we assayed *S. paradoxus* Z1, *S. cerevisiae* DBVPG1373, and the full *ESP1* swap in the *S. cerevisiae* background (harboring the promoter and open reading frame from *S. paradoxus*) from (Weiss et al., 2018a) as follows. Each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at room temperature for 3 days. For each biological replicate, a single colony was inoculated into 5 mL liquid YPD and grown for 48 hours at 28°C with shaking at 200 rpm to create a stationary phase pre-culture. From pre-culture we made eight back-dilution experimental cultures in a standard PCR strip tube, each in 200 µL YPD, and we incubated these in a thermocycler using a gradient protocol from 37.0 to 40.8°C. After 24 hours, 150 µL from each culture was removed and OD_{600} was measured. The pipeline from inoculation off of solid plates through pre-culture, back-dilution, and growth we refer to as a day's growth experiment for the dose-response of a strain. For each day's experiments, at a given temperature we calculated the average efficiency $\langle e_{Scer,37} \rangle$ across the replicates of wild-type *S. cerevisiae* at 37°C, and used it to normalize the efficiency $e_{s,T}$ measured for each replicate assayed on that day of a strain of interest *s* at temperature T. Thus, the final measurement used for analysis for each replicate and temperature on a given day was $e_{s,T}/\langle e_{Scer,37} \rangle$. We carried out two days' worth of replicate growth experiments, and used the complete cohort of measurements of $e_{s,T}/\langle e_{Scer,37} \rangle$ from all days and all temperatures as input into a two-factor type 2 ANOVA test for a temperature-by-strain effect comparing *s* with *S. cerevisiae*.

Microscopy

Microscopy was performed as described in (Weiss et al., 2018a). Images were scored, blinded, for the size of dyads, omitting all clumps of >2 cells. Two replicates per strain and condition were imaged, and ten to sixteen images per replicate were scored. Significance was evaluated using a two-factor ANOVA test to evaluate strain by temperature effects. The range and mean number of dyads scored per image and per strain are reported in Table S7.

RESULTS

Signatures of adaptation and constraint at *S. cerevisiae* thermotolerance loci

With the goal of investigating evolutionary mechanisms of thermotolerance divergence between yeasts, we started by addressing the genetics of the trait. Our earlier study used genome-scale screens with the reciprocal hemizygosity test (Steinmetz et al., 2002; Stern, 2014) to identify eight genes at which *S. cerevisiae* harbored pro-thermotolerance alleles relative to those of *S. paradoxus* (Weiss et al., 2018a). We re-processed these screen data with an improved statistical workflow to boost power and genome coverage (see Methods). The results recapitulated seven loci that we had reported and validated, plus an additional seven that had not risen to significance in our original analysis (Table S1). We considered the expanded set of loci as a more complete model of the genetic architecture of the trait, which would be well-suited to population and evolutionary analyses.

Thermotolerance is a defining and putatively adaptive character of *S. cerevisiae*, shared among isolates within the species and distinguishing it from the rest of the *Saccharomyces* clade (Gonçalves et al., 2011; Salvadó et al., 2011; Sweeney et al., 2004). We hypothesized that the loci underlying thermotolerance had evolved under positive selection before the radiation of modern *S. cerevisiae* populations. To test this, we made use of a broad population survey of *S. cerevisiae* (Peter et al., 2018), and the deepest-sampled *S. paradoxus* populations available (from vineyards and European collection locales (Bergström et al., 2014) and from North America (Durand et al., 2019)). With these genomes, we first sought to quantify sequence diversity between the species, at thermotolerance genes. The absolute diversity statistic D_{xy} reaches high levels in a lineage after selection when compared to a representative of the ancestral state (Nei, 1987), and is preferred over relative-divergence metrics as a suggestive statistic of adaptation (Noor and Bennett, 2009). Using the entire set of population genomes from *S. cerevisiae* and S. *paradoxus*, we found enrichment for high D_{xy} among our thermotolerance genes (Table 1 and Table S2), as expected from a previous smaller-scale analysis (Weiss et al., 2018a). The latter result was mirrored by analyses of individual *S. cerevisiae* populations (Table S3), ruling out demographic artifacts as the source of signal in our species-wide test. Thus, divergence from *S. paradoxus* at thermotolerance loci is a trend that pervades >30 *S. cerevisiae* populations, collected in Europe, Asia, Africa, and the Americas, supporting a model of a selective event in the ancestor of modern *S. cerevisiae*.

We next reasoned that, if evolution had used predominantly amino acid variants in building the thermotolerance trait, the underlying loci would exhibit striking coding variation between species, relative to within-species polymorphism and relative to synonymous changes, as analyzed in the family of methods derived from the McDonald-Kreitman test (McDonald and Kreitman, 1991). For enrichment tests we used the neutrality index (Stoletzki and Eyre-Walker, 2011), which reaches low values in cases of adaptive amino-acid evolution between species. The results revealed a 1.37-fold reduced neutrality index among thermotolerance genes relative to the genome as a

whole (Table 1 and Table S2). Such a signal strongly supports a history of adaptation at thermotolerance loci, with a mechanism involving changes to protein structure and/or function.

Under our model of thermotolerance evolution, after an ancestral *S. cerevisiae* population gained the trait long ago, it was maintained by purifying selection throughout the species. To assess signatures of constraint within *S. cerevisiae* on thermotolerance loci, we reasoned that haplotype-level analyses would have greater power than site-by-site tests. The SelectionHapStats suite (Garud et al., 2015; Harris et al., 2018) can be used for this purpose to detect very recent soft selective sweeps in young populations or, for a putatively ancient adaptation like ours, to report conservation more generally. In analyses using our complete set of genomes from *S. cerevisiae*, thermotolerance loci were enriched for high genotype homozygosity (Table 1 and Table S2), as seen at any given selected site after a sweep as a product of strong allele-sharing (Garud et al., 2015; Harris et al., 2018). As a control for potential demographic effects in this whole-species analysis, we repeated the test paradigm on individual well-sampled *S. cerevisiae* populations, and again detected elevated genotype homozygosity at thermotolerance genes (Tables S2 and S4). Together, our sequence-based analyses establish hallmarks of directional selection at these loci: sequence divergence between *S. cerevisiae* and *S. paradoxus*, particularly at amino-acid coding sites, and tight constraint within *S. cerevisiae*.

Molecular evolution and functional impact of coding variation at *ESP1*

We anticipated that inspecting allele-sharing within species, at high resolution across genomic loci, could further help reveal facets of the history of thermotolerance genes. Using the largest well-sampled population of *S. cerevisiae* (collected from localities across Europe and in vineyards elsewhere), we found that genotype homozygosity was not uniform across a given thermotolerance gene, and for most loci, peaks of allele-sharing could be resolved (Figure 1A). Notably, even at allele-sharing peaks, in absolute terms genotype homozygosity was modest. Across all thermotolerance loci, the top-scoring regions in the wine/European population was at the 5' end of the chromosome segregation gene *ESP1* (Figure 1A), where the statistic reached at most a value of 0.15. This is consistent with our inference of an ancient date for positive selection at *ESP1* and other loci, since the tight conservation and long haplotypes expected immediately after selection would be eroded over longer timescales (Berry et al., 1991; Smith and Haigh, 1974; Weigand and Leese, 2018). Using this very highest region of allele-sharing in *ESP1* as a test case, we inspected it in other well-sampled *S. cerevisiae* populations and again found elevated genotype homozygosity (Figure 1B-E), indicating that all these populations likely have had the same forces at play at the locus.

We next sought to gain deeper molecular insight into thermotolerance genetics and, for this purpose, chose to focus further on *ESP1* as a testbed, given its high allele-sharing within *S. cerevisiae* (Figure 1 and Table S2) and dramatic impact on the thermotolerance trait (Weiss et al., 2018a). First, using a phylogenetic approach across the *Saccharomyces* genus, we established a pattern of accelerated protein evolutionary

rate along the *S. cerevisiae* lineage in *ESP1* ($p = 0.046$), consistent with our population-level tests of protein evolution on the larger set of thermotolerance genes (Table 1). Next, to investigate the importance of coding variation at *ESP1* experimentally, we turned to an allele-swap design. We introduced the *ESP1* coding region and, separately, the *ESP1* promoter, from wild *S. paradoxus* (strain Z1, isolated from an oak tree in the United Kingdom) into a wild *S. cerevisiae* background (strain DVBPG1373, from soil in the Netherlands). Growth experiments revealed a dramatic, temperature-dependent effect of variation in the *ESP1* coding region, with the *S. paradoxus* allele compromising growth under heat treatment (Figure 2 and Figure S2). This transgenic fully recapitulated the impact of a larger, regional swap of the *S. paradoxus* open reading frame and promoter together into *S. cerevisiae* (Figure 2 and Figure S2). By contrast, the *S. paradoxus* allele of the *ESP1* promoter conferred no defect in thermotolerance when analyzed on its own (Figure 2 and Figure S2). As an independent analysis of potential promoter effects, we examined *cis*-regulatory variation between yeast species in measurements of *ESP1* gene expression. We found no overall dramatic tendency for overall *cis*-regulatory divergence between *S. cerevisiae* and other species, at *ESP1* in particular or across thermotolerance genes as a set (Table S5A). Likewise, the latter yielded no signal in tests for directional cis-regulatory divergence ((Bullard et al., 2010); Table S5B). Together, our data highlight the evolutionary and functional importance of amino acid variation between *S. cerevisiae* and *S. paradoxus* at *ESP1,* and raise the possibility that coding divergence may also underlie the thermotolerance effects of other mapped loci.

Temperature dependence and cell biology of species divergence effects at *ESP1*

In further pursuit of the molecular mechanisms of *S. cerevisiae* thermotolerance, we turned to the potential for clues from temperature-dependent genetics. *S. cerevisiae* outperforms its sister species at a range of elevated temperatures (Salvadó et al., 2011; Sweeney et al., 2004). Our thermotolerance loci were identified in a screen for effects of interspecies divergence at 39°C (Weiss et al., 2018a), and their relevance to growth under other conditions is unknown. Drawing again on *ESP1* as a model with which to address this question, we assayed biomass accumulation of wild-type and transgenic strains under a temperature dose-response. In these growth experiments, we observed a gradual decline in wild-type *S. cerevisiae* and *S. paradoxus* growth as temperature increased, with the latter more sensitive to heat as expected (Figure 3; (AlZaben et al., 2021). Our allele-swap strain in the *S. cerevisiae* background harboring *S. paradoxus ESP1* exhibited a sharp drop in growth at ~38°C; it grew readily below this temperature, phenocopying the wild-type *S. cerevisiae* progenitor, and at higher temperatures, it exhibited the negligible growth seen in wild-type *S. paradoxus* (Figure 3). Such a dose-response, resembling the sigmoidal behavior of a cooperative biochemical process, was a synthetic property of the *ESP1* inter-species swap, distinguishing it from either wild-type species. These data imply that, at least in the *S. cerevisiae* background, the function of *S. paradoxus* Esp1 breaks down with a steep temperature dependence, whose midpoint is close to the conditions under which this gene was originally identified (39°C).

ESP1 encodes separase, which releases sister chromatids for separation into daughter cells during anaphase, cleaving the cohesin ring that has held them together in metaphase. We reasoned that, if *S. paradoxus* Esp1 failed to function, in actively growing cells harboring this allele we would see hallmarks of arrest late in the cell cycle. Quantitative microscopy bore out this prediction: as in wild-type *S. paradoxus* (Weiss et al., 2018a), large-budded dyads predominated in cultures of the *S. cerevisiae* transgenic with *S. paradoxus ESP1*, when incubated at 39°C (Figure 4). These findings are consistent with a mechanism in which heat treatment compromises separase function of the *S. paradoxus* allele of Esp1, blocking the progress of the cell cycle and limiting viability and biomass accumulation. Under such a model, evolution in *S. cerevisiae* would have resolved these defects, introducing genetic changes that foster Esp1 function and boost fitness at high temperature.

DISCUSSION

In the study of adaptation, a trait that arises in a species, goes to fixation, and is maintained for thousands of generations can be seen as the ultimate evolutionary success story. Here we have used yeast thermotolerance as a model of this process. We shed light on the forces driving the trait as it has evolved between species, and we investigated the molecular genetics and cell biology of divergent alleles at the underlying loci.

Our sequence-based tests of thermotolerance loci—revealing divergence and protein evolutionary rate between species, and conservation within *S. cerevisiae*—strongly suggest that the trait arose under a selective sweep before the radiation of modern *S. cerevisiae* populations. Also consistent with the latter model is the fact that at a given thermotolerance gene, alleles from *S. cerevisiae* isolates from around the world were partially sufficient for the trait, when swapped into a poorly-performing *S. paradoxus* background (Weiss et al., 2018a). Plausibly, the initial rise of thermotolerance early in *S. cerevisiae* history could have been driven by the ecology of hot East Asian niches where the species likely originated (Peter et al., 2018).

The scenario of an ancient sweep of thermotolerance, whose effects bear out across modern *S. cerevisiae*, sets up an intriguing contrast with traits that undergo independent, parallel adaptations in distinct lineages of a species (Chan et al., 2010; Hoekstra and Nachman, 2003; Rosenblum et al., 2010; Xie et al., 2019). Under one compelling model, thermotolerance alleles acquired by an initially small, specialized *S. cerevisiae* ancestor could have enabled later migrants to colonize other warm niches (Robinson et al., 2016). That said, additional lineage-specific adaptations to heighten thermotolerance further could also eventually come to light.

Our data also open a window onto the molecular mechanisms of thermotolerance evolution in the yeast system. The patterns we have seen of non-synonymous sequence variation in thermotolerance genes, and our molecular-genetic experiments at the *ESP1* locus, point to a key role for protein-coding variation. What exactly would such amino acid changes be doing at thermotolerance loci? Given how much our

temperature dose-response of *ESP1* function looks like a two-state protein unfolding curve, it is tempting to speculate that the *S. cerevisiae* allele of this gene may act by boosting protein stability. If such a mechanism were to prove the general rule for our loci, it would dovetail with the trend for proteome thermostability seen in heat-tolerant species (Leuenberger et al., 2017). Perhaps most likely, however, given the complexity of the trait, is a picture in which evolution tweaked many protein features (*e.g.* (Sas-Chen et al., 2020)), at different times and at different loci, as the *S. cerevisiae* ancestor gained its unique thermotolerance character.

DATA AVAILABILITY

Strains and plasmids are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables. Custom scripts for sequence preparation and population genetics statistics for absolute sequence divergence, absolute nucleotide divergence, and the McDonald-Kreitman neutrality index are available at https://github.com/clairedubin/thermotolerance. Custom scripts for RH-seq reanalysis, multi-locus genotype and allele-sharing inference, and allele-specific expression analysis are available at https://github.com/melanieabrams-pub/thermotolerance-loci-across-yeasts.

ACKNOWLEDGEMENTS

The authors thank Julie Chuong for verifying and maintaining transgenic strains, David Savage for his generosity with lab facilities and resources, Josh Schraiber for discussions, and Jeremy Roop for critical reading of the manuscript. This work was supported by NSF GRFP DGE 1752814 to M.A. and NIH R01 GM120430 to R.B.B.

REFERENCES

AlZaben, F., Chuong, J.N., Abrams, M.B., and Brem, R.B. (2021). Joint effects of genes underlying a temperature specialization tradeoff in yeast. PLoS Genet 17, e1009793.

Anderson, D.P., Whitney, D.S., Hanson-Smith, V., Woznica, A., Campodonico-Burnett, W., Volkman, B.F., King, N., Thornton, J.W., and Prehoda, K.E. (2016). Evolution of an ancient protein function involved in organized multicellularity in animals. ELife 5, e10147.

Anderson, D.W., McKeown, A.N., and Thornton, J.W. (2015). Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. ELife 4, e07864.

Asgari, S., Luo, Y., Akbari, A., Belbin, G.M., Li, X., Harris, D.N., Selig, M., Bartell, E., Calderon, R., Slowikowski, K., et al. (2020). A positively selected FBN1 missense variant reduces height in Peruvian individuals. Nature 582, 234–239.

Baldwin, M.W., Toda, Y., Nakagita, T., O'Connell, M.J., Klasing, K.C., Misaka, T., Edwards, S.V., and Liberles, S.D. (2014). Sensory biology. Evolution of sweet

taste perception in hummingbirds by transformation of the ancestral umami receptor. Science 345, 929–933.
Barroso-Batista, J., Sousa, A., Lourenço, M., Bergman, M.-L., Sobral, D., Demengeot, J., Xavier, K.B., and Gordo, I. (2014). The first steps of adaptation of Escherichia coli to the gut are dominated by soft sweeps. PLoS Genetics 10, e1004182.
Bergström, A., Simpson, J.T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A.N., Moses, A.M., Louis, E.J., Mustonen, V., et al. (2014). A high-definition view of functional genetic variation from natural yeast genomes. Mol. Biol. Evol. 31, 872–888.
Berry, A.J., Ajioka, J.W., and Kreitman, M. (1991). Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics 129, 1111–1117.
Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012). Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 489, 513–518.
Bridgham, J.T., Ortlund, E.A., and Thornton, J.W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461, 515–519.
Bullard, J.H., Mostovoy, Y., Dudoit, S., and Brem, R.B. (2010). Polygenic and directional regulatory evolution across pathways in Saccharomyces. Proc. Natl. Acad. Sci. U.S.A. 107, 5058–5063.
Castro, J.P., Yancoskie, M.N., Marchini, M., Belohlavy, S., Hiramatsu, L., Kučka, M., Beluch, W.H., Naumann, R., Skuplik, I., Cobb, J., et al. (2019). An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. ELife 8, e42014.
Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science (New York, N.Y.) 327, 302–305.
Cleves, P.A., Ellis, N.A., Jimenez, M.T., Nunez, S.M., Schluter, D., Kingsley, D.M., and Miller, C.T. (2014). Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. PNAS 111, 13912–13917.
Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422–1423.
Durand, É., Gagnon-Arsenault, I., Hallin, J., Hatin, I., Dubé, A.K., Nielly-Thibault, L., Namy, O., and Landry, C.R. (2019). Turnover of ribosome-associated transcripts from de novo ORFs produces gene-like characteristics available for de novo gene emergence in wild yeast populations. Genome Res. 29, 932–943.
Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Field, Y., Boyle, E.A., Telis, N., Gao, Z., Gaulton, K.J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M.I., et al. (2016). Detection of human adaptation during the past 2000 years. Science 354, 760–764.
Finnigan, G.C., Hanson-Smith, V., Stevens, T.H., and Thornton, J.W. (2012). Evolution of increased complexity in a molecular machine. Nature 481, 360–364.

Garud, N.R., Messer, P.W., Buzbas, E.O., and Petrov, D.A. (2015). Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. PLoS Genet 11, e1005004.
Gonçalves, P., Valério, E., Correia, C., Almeida, J.M.G.C.F. de, and Sampaio, J.P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric Saccharomyces Species. PLOS ONE 6, e20739.
Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., and Desai, M.M. (2017). The dynamics of molecular evolution over 60,000 generations. Nature 551, 45–50.
Harris, A.M., Garud, N.R., and DeGiorgio, M. (2018). Detection and Classification of Hard and Soft Sweeps from Unphased Genotypes by Multilocus Genotype Identity. Genetics 210, 1429–1452.
Hoekstra, H.E., and Nachman, M.W. (2003). Different genes underlie adaptive melanism in different populations of rock pocket mice. Molecular Ecology 12, 1185–1194.
Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P.J., von Mering, C., Claassen, M., and Picotti, P. (2017). Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science 355.
Li, X.C., and Fay, J.C. (2017). Cis-Regulatory Divergence in Gene Expression between Two Thermally Divergent Yeast Species. Genome Biol Evol 9, 1120–1129.
Linnen, C.R., Poh, Y.-P., Peterson, B.K., Barrett, R.D.H., Larson, J.G., Jensen, J.D., and Hoekstra, H.E. (2013). Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. Science 339, 1312–1316.
Liu, Q., Onal, P., Datta, R.R., Rogers, J.M., Schmidt-Ott, U., Bulyk, M.L., Small, S., and Thornton, J.W. (2018). Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. Elife 7.
Massey, J., and Wittkopp, P.J. (2016). The genetic basis of pigmentation differences within and between Drosophila species. Curr Top Dev Biol 119, 27–61.
McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654.
Nei, M. (1987). Molecular Evolutionary Genetics (Columbia University Press).
Noor, M. a. F., and Bennett, S.M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb) 103, 439–444.
Peter, J., Chiara, M.D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344.
Pillai, A.S., Chandler, S.A., Liu, Y., Signore, A.V., Cortez-Romero, C.R., Benesch, J.L.P., Laganowsky, A., Storz, J.F., Hochberg, G.K.A., and Thornton, J.W. (2020). Origin of complexity in haemoglobin evolution. Nature 581, 480–485.
Robinson, H.A., Pinharanda, A., and Bensasson, D. (2016). Summer temperature can predict the distribution of wild yeast populations. Ecology and Evolution 6, 1236–1250.
Rosenblum, E.B., Römpler, H., Schöneberg, T., and Hoekstra, H.E. (2010). Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proceedings of the National Academy of Sciences of the United States of America 107, 2113–2117.

Sackton, T.B., Grayson, P., Cloutier, A., Hu, Z., Liu, J.S., Wheeler, N.E., Gardner, P.P., Clarke, J.A., Baker, A.J., Clamp, M., et al. (2019). Convergent regulatory evolution and loss of flight in paleognathous birds. Science (New York, N.Y.) 364, 74–78.

Salvadó, Z., Arroyo-López, F.N., Guillamón, J.M., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus Saccharomyces. Appl Environ Microbiol 77, 2292–2302.

Sas-Chen, A., Thomas, J.M., Matzov, D., Taoka, M., Nance, K.D., Nir, R., Bryson, K.M., Shachar, R., Liman, G.L.S., Burkhart, B.W., et al. (2020). Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. Nature 583, 638–643.

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genetical Research 23, 23–35.

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

Stern, D.L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. Trends Genet 30, 547–554.

Stoletzki, N., and Eyre-Walker, A. (2011). Estimation of the neutrality index. Mol. Biol. Evol. 28, 63–70.

Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D., and Lynch, V.J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. ELife 5, e11994.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34, W609-612.

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

Tenaillon, O., Barrick, J.E., Ribeck, N., Deatherage, D.E., Blanchard, J.L., Dasgupta, A., Wu, G.C., Wielgoss, S., Cruveiller, S., Médigue, C., et al. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. Nature 536, 165–170.

Tian, X., Firsanov, D., Zhang, Z., Cheng, Y., Luo, L., Tombline, G., Tan, R., Simon, M., Henderson, S., Steffan, J., et al. (2019). SIRT6 Is Responsible for More Efficient DNA Double-Strand Break Repair in Long-Lived Species. Cell 177, 622-638.e22.

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

Toprak, E., Veres, A., Michel, J.-B., Chait, R., Hartl, D.L., and Kishony, R. (2011). Evolutionary paths to antibiotic resistance under dynamically sustained drug stress. Nat Genet 44, 101–105.

Weigand, H., and Leese, F. (2018). Detecting signatures of positive selection in non-model species using genomic data. Zool J Linn Soc 184, 528–583.

Weiss, C.V., Roop, J.I., Hackley, R.K., Chuong, J.N., Grigoriev, I.V., Arkin, A.P., Skerker, J.M., and Brem, R.B. (2018). Genetic dissection of interspecific differences in yeast thermotolerance. Nature Genetics 50, 1501.

Will, J.L., Kim, H.S., Clarke, J., Painter, J.C., Fay, J.C., and Gasch, A.P. (2010). Incipient balancing selection through adaptive loss of aquaporins in natural Saccharomyces cerevisiae populations. PLoS Genet. 6, e1000893.

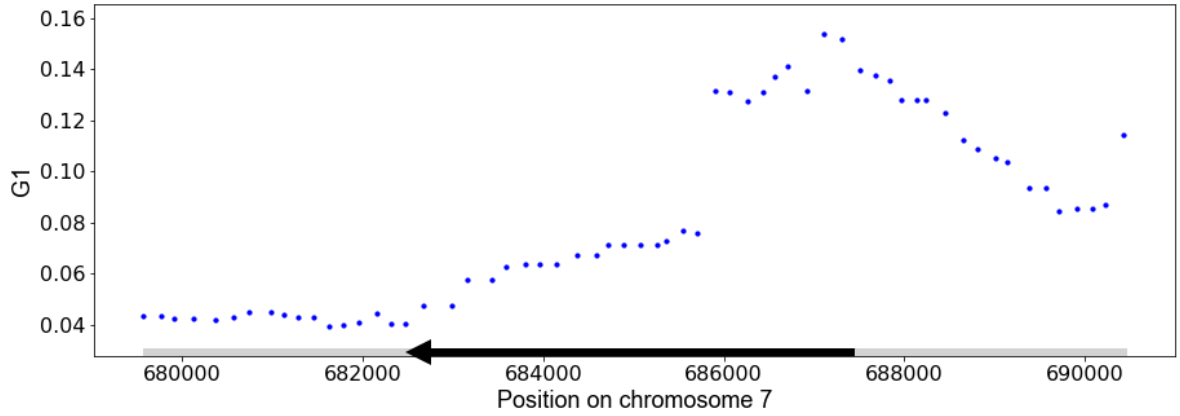
Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science 285, 901–906.

Xie, K.T., Wang, G., Thompson, A.C., Wucherpfennig, J.I., Reimchen, T.E., MacColl, A.D.C., Schluter, D., Bell, M.A., Vasquez, K.M., and Kingsley, D.M. (2019). DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. Science 363, 81–84.

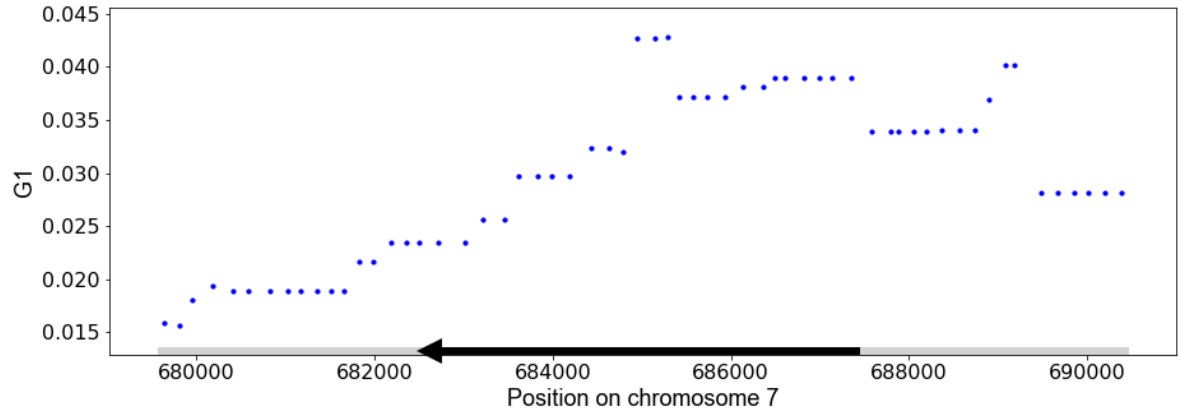
Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24, 1586–1591.

FIGURES

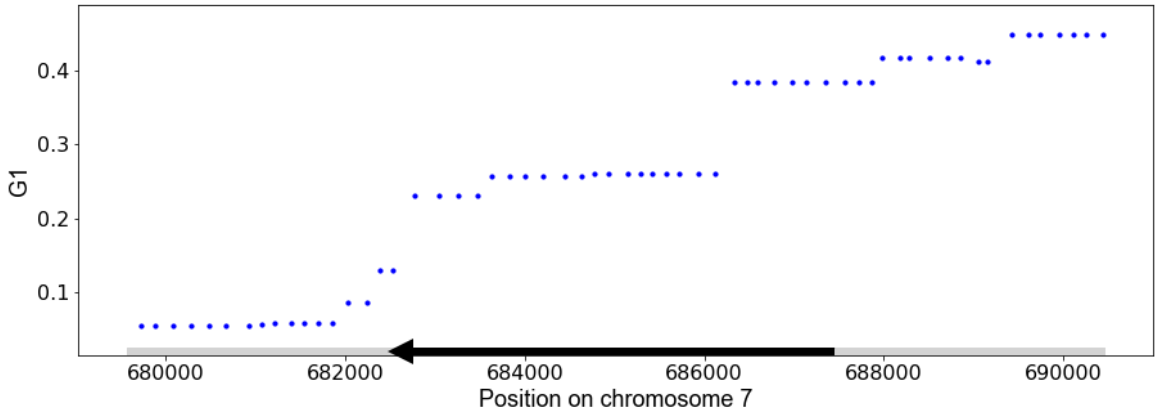
A.



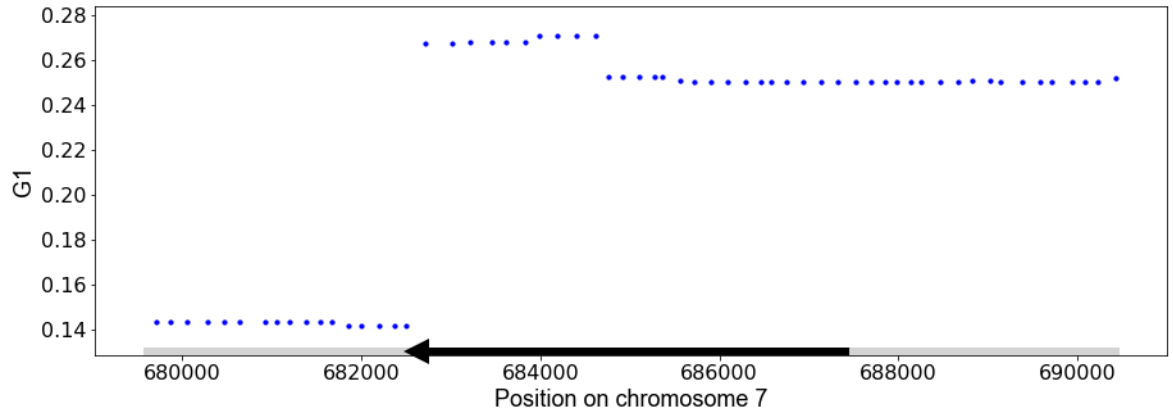
B.



C.



D.



E.

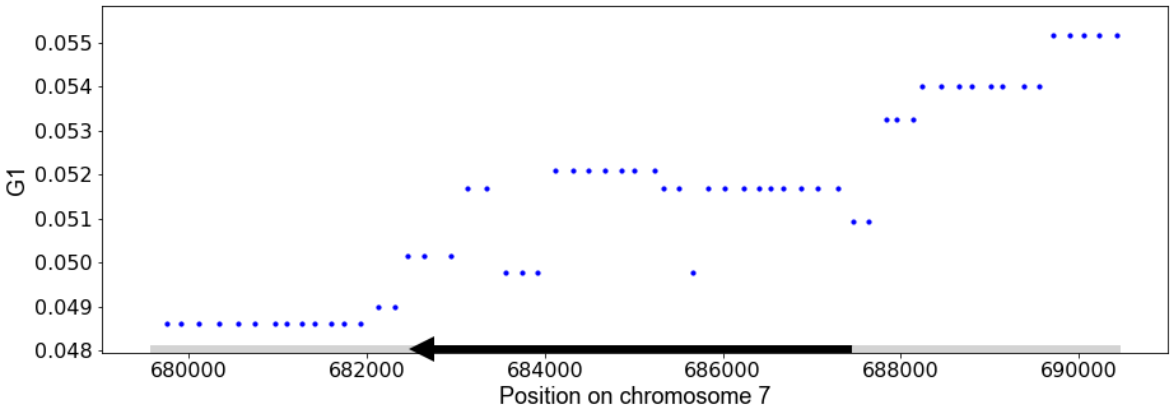


Figure 1. A peak of high allele frequency in *S. cerevisiae* populations at the 5' end of *ESP1*. Each panel shows results of analysis of allele frequency at the thermotolerance gene *ESP1* in a population of *S. cerevisiae* from (Peter et al., 2018). In each panel, the *y*-axis reports genotype homozygosity, G1, in a 1200-SNP window around the position shown on the *x*. The *ESP1* open reading frame is demarcated with a dark black arrow (direction of transcription is right to left). **A**, Wine/European population. **B**, Mosaic Region 3 population. **C**, Brazilian Bioethanol population**. D**, Sake population. **E**, Mixed Origin population.

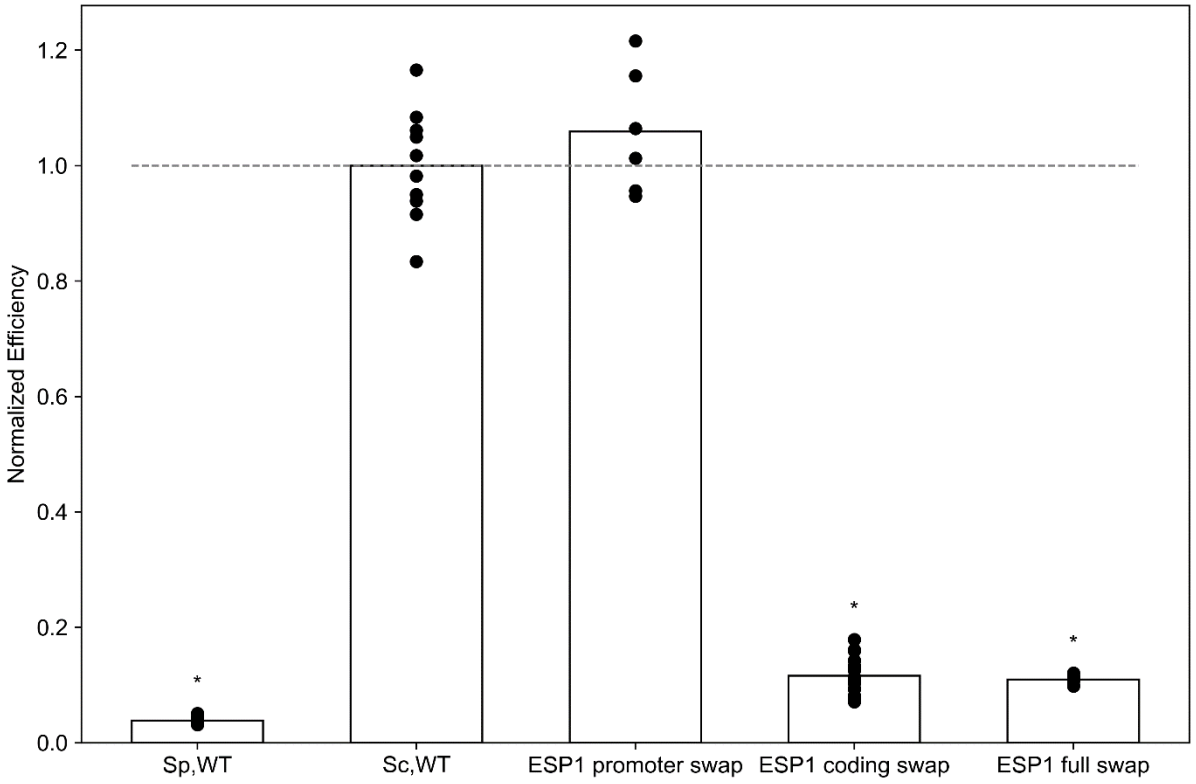


Figure 2. The *S. cerevisiae ESP1* coding region, but not the promoter, is required for thermotolerance. Each column represents results from biomass accumulation assays of a wild-type or transgenic yeast strain cultured at high temperature. The *y*-axis reports the optical density of a culture of the indicated strain after 24h at 39°C, normalized to the analogous quantity from wild-type *S. cerevisiae* (dashed line). Each point reports results from one biological replicate, and each bar height reports the average across replicates (n = 6-18). The first two columns report results from wild-type (WT) strains of *S. paradoxus* Z1 (Sp) and *S. cerevisiae* DBVPG17373 (Sc). The last three columns report strains with the indicated region of *ESP1* from *S. paradoxus* swapped into *S. cerevisiae* at the endogenous location; *ESP1* full swap denotes transgenesis of both the promoter and the coding region. *, Wilcoxon test $p < 0.004$ in a comparison against wild-type *S. cerevisiae*. Culture data at 28°C are given in Figure S2.

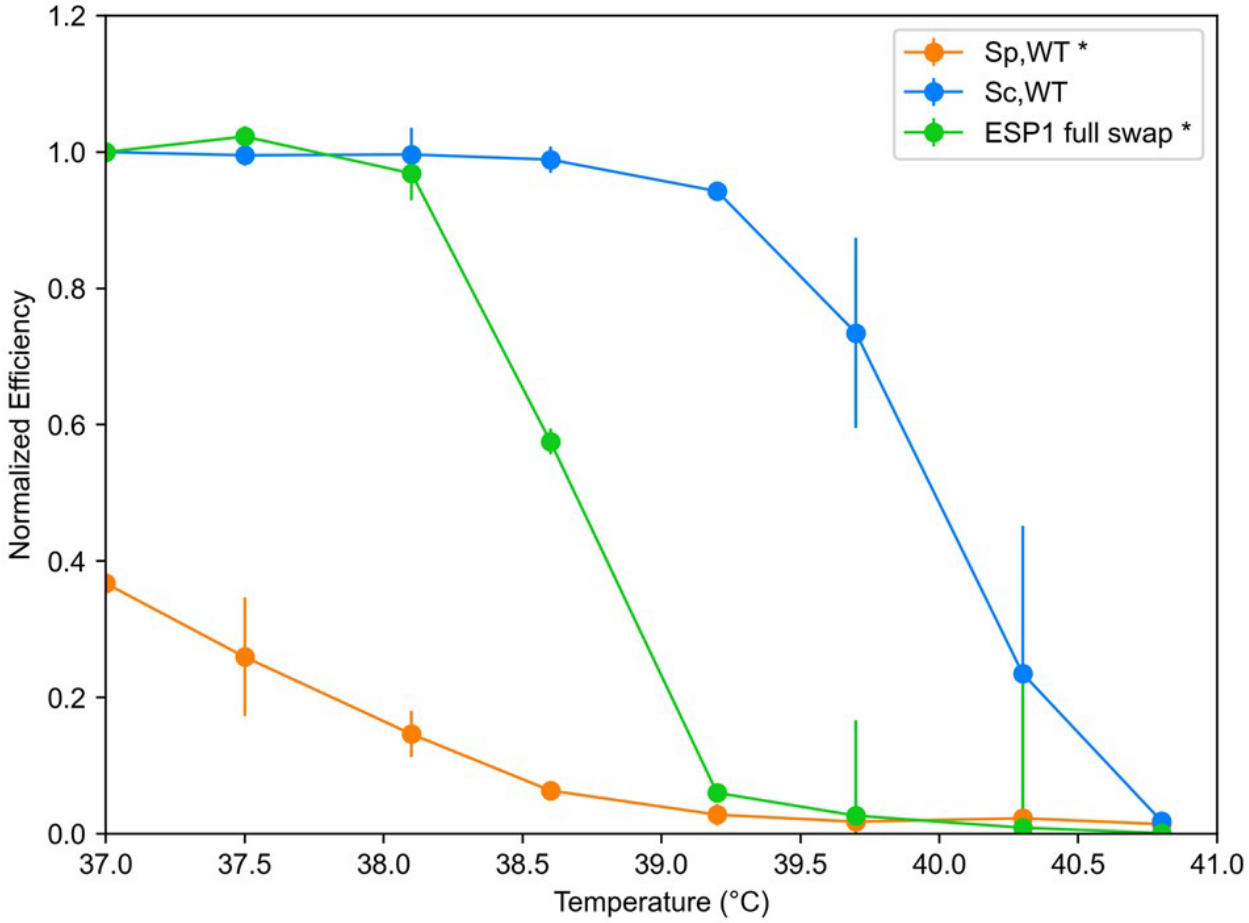
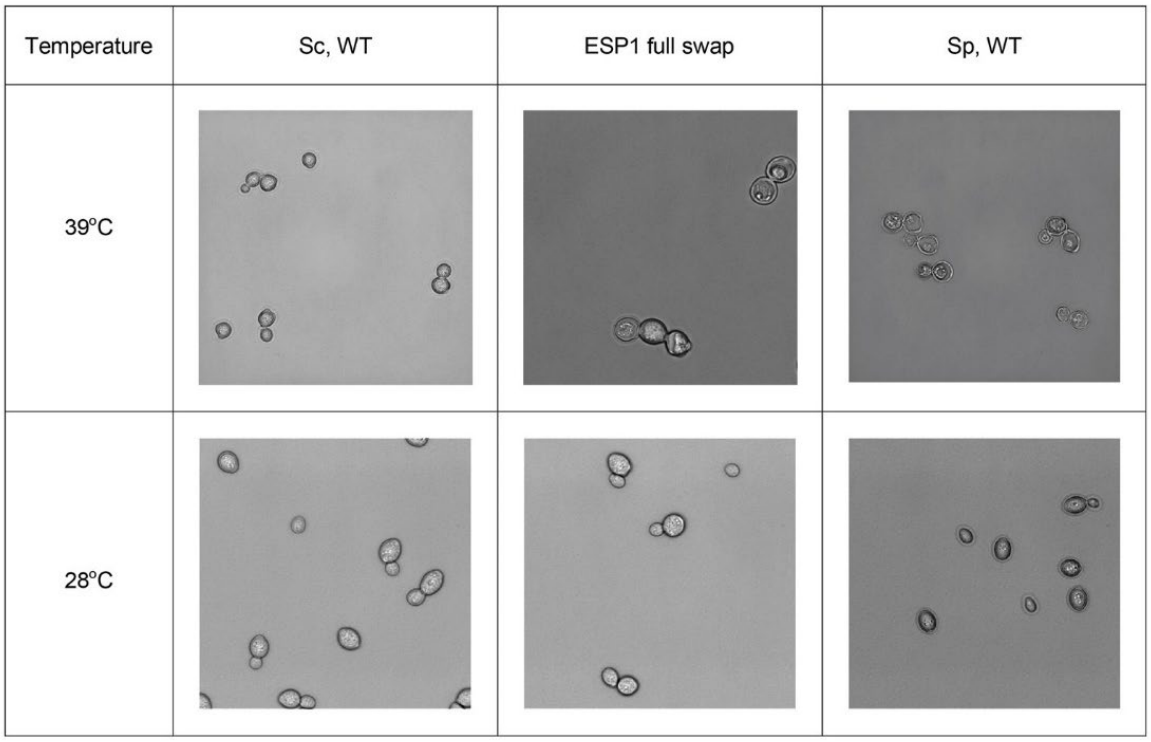
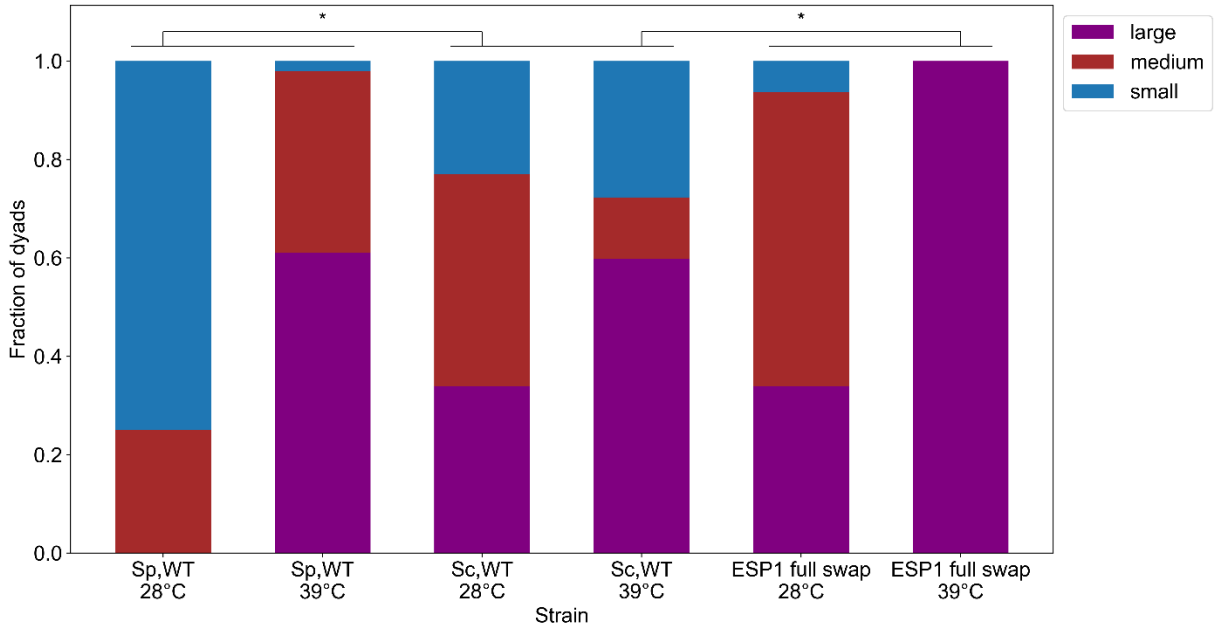


Figure 3. Growth function of *S. paradoxus ESP1* declines sharply with temperature. Each trace reports results from biomass accumulation assays of a wild-type or transgenic yeast strain across temperatures. Strain labels are as in Figure 2. The *y*-axis reports the optical density of a culture of the indicated strain after 24h at the temperature on the *x*, normalized to the optical density of that day's wild-type *S. cerevisiae* at 37°C. *, $p < 10^{-12}$ for the strain by temperature interaction term of a two-factor ANOVA, in a comparison between the indicated strain and wild-type *S. cerevisiae*.

A.

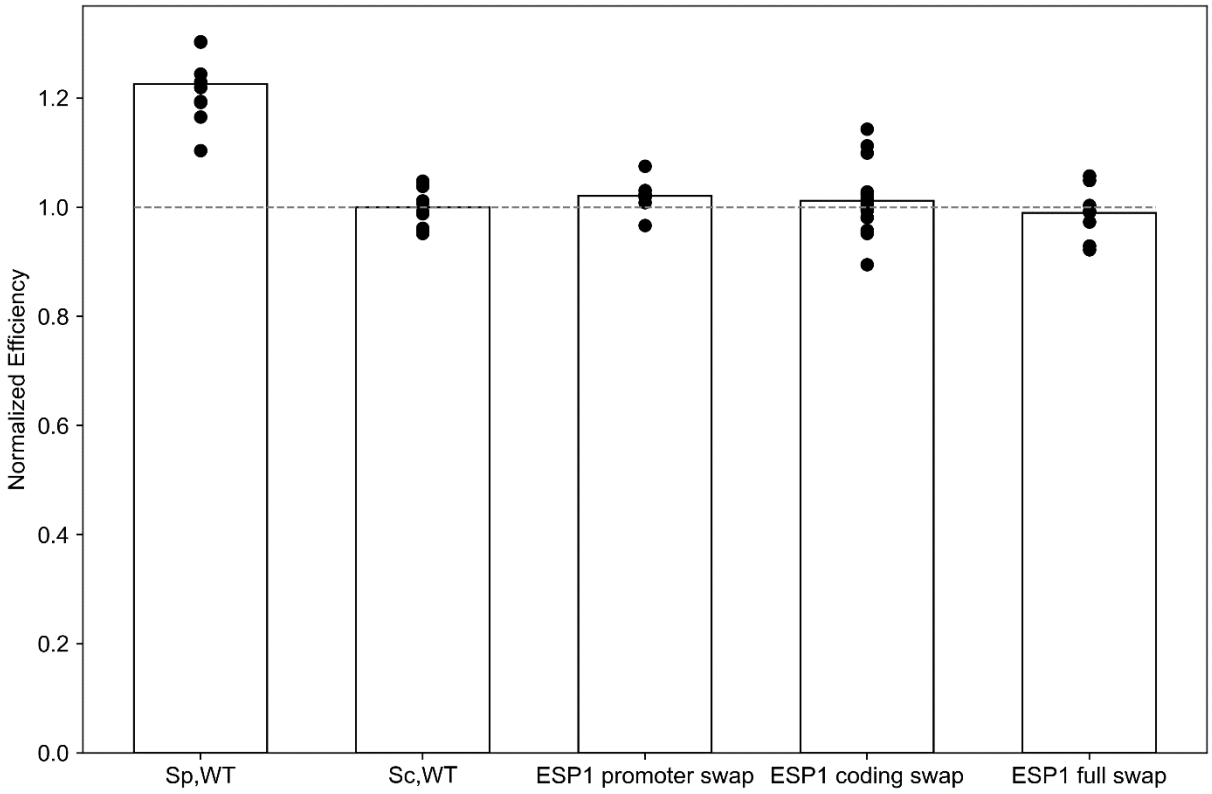


B.



Each panel reports a representative image of a wild-type or transgenic yeast strain after incubation for 24h at the indicated temperature. Strain labels are as in Figure 2. **B**, Each bar reports quantification of replicated imaging data of the indicated

strain cultured at the indicated temperature, as in **A**. For each bar, the *y*-axis shows the fraction of dyads in the indicated size category. *, $p < 0.015$ for the strain by temperature interaction term of a two-factor ANOVA, in a comparison between the indicated strain and wild-type *S. cerevisiae*. Experiment details are given in Table S7.


Normalized Efficiency
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Sp,WT
Sc,WT
ESP1 promoter swap
ESP1 coding swap
ESP1 full swap

TABLES

	Thermotolerance genes	Genome	*p*
D_{xy}[a]	0.1215	0.1016	0.0053*
NI[b]	2.0513	2.8076	0.0217*
G1[c]	0.01025	0.0077	0.0491*

Table 1. Thermotolerance loci are enriched for positive selection between species and allele-sharing in *S. cerevisiae*. [a]The columns report the median value of the absolute diversity statistic D_{xy} in thermotolerance genes and random sets of genes from the genome, respectively, and empirical significance from a resampling test. [b]Data are as in ***a*** except that the metric analyzed was the neutrality index NI according to the McDonald-Kreitman test. [c]Data are as in ***a*** except that the metric analyzed was the allele-sharing statistic genotype homozygosity, G1. *, $p < 0.05$.

gene	p (Weiss et al, 2018)[a]	p (revised pipeline)[b]
YLR397C/AFG2	7.0×10^{-4}	1.81×10^{-7}
YGR098C/ESP1	3.8×10^{-10}	4.65×10^{-7}
YMR168C/CEP3	5.4×10^{-3}	9.04×10^{-6}
YKR054C/DYN1	1.6×10^{-6}	5.16×10^{-5}
YHR023W/MYO1	2.8×10^{-6}	1.09×10^{-4}
YDR180W/SCC2	7.8×10^{-3}	7.21×10^{-4}
YPL174/NIP100C	n.s.	3.90×10^{-3}
YPR164W/MMS1	n.s.	4.06×10^{-3}
YCR042C/TAF2	3.1×10^{-3}	4.79×10^{-3}
YMR016C/SOK2	n.s.	0.01883
YJR135C/MCM22	n.s.	0.02596
YJL025W/RRN7	n.s.	0.02858
YDR443C/SSN2	n.s.	0.03813
YKL134C/OCT1	n.s.	0.04997
# genes tested	3416	4580
fraction of genes tested	0.54	0.72

Supplementary Table 1. Re-analysis of a screen for genes at which divergence between *S. cerevisiae* and *S. paradoxus* contributes to high-temperature growth.
[a]Published results from a reciprocal hemizygosity (RH) screen for thermotolerance in a *S. cerevisiae* x *S. paradoxus* hybrid, using an analysis scheme in which the average fitness of a given mutant across biological replicates is used as input into the RH test (Weiss et al., 2018). The first ten rows report significance of the indicated gene in the RH test (comparing thermotolerance of two sets of hemizygotes, bearing disruptions in the two species' alleles respectively, with correction for multiple testing). The bottom two rows report the number and fraction of genes tested in this pipeline. **[b]**Results are as in ***a*** except that a distinct analysis pipeline was used, in which fitness measurements for a given mutant from biological replicates are used as independent input into the RH test. n.s., not significant at corrected $p < 0.05$.

gene	D_{xy}	NI	G1
YLR397C/AFG2	0.1047	2.0332	0.0089
YGR098C/ESP1	0.1194	1.9854	0.0157
YMR168C/CEP3	N/A	N/A	0.0093
YKR054C/DYN1	0.1251	1.9515	0.0147
YHR023W/MYO1	0.1156	2.7887	0.0135
YDR180W/SCC2	0.1215	1.9801	0.0134
YPL174/NIP100C	0.1549	1.4479	0.0190
YCR042C/TAF2	0.1326	2.4492	0.0103
YMR016C/SOK2	0.12	2.0515	0.0108
YJR135C/MCM22	0.1419	1.8593	0.0064
YJL025W/RRN7	0.1229	4.1086	0.0093
YDR443C/SSN2	0.0992	2.738	0.0054
YKL134C/OCT1	0.0875	3.9397	0.0056

Supplementary Table 2. Metrics of interspecies diversity and intraspecies polymorphism at thermotolerance genes. Each row reports, for the indicated thermotolerance gene, the absolute diversity statistic D_{xy}, the neutrality index NI according to the McDonald-Kreitman test, and the genotype homozygosity, G1.

A.

S. cerevisiae population	Thermotolerance genes[a]	Genome[b]	p-value[c]
French Guiana Human	0.1213	0.1016	0.0044
Ale Beer	0.1185	0.0996	0.0043
West African Cocoa	0.1203	0.1003	0.0036
African Palm Wine	0.1210	0.1010	0.0035
CHNIII	0.1206	0.1014	0.0034
CHNII	0.1199	0.1003	0.0036
CHNI	0.1214	0.1012	0.0031
Taiwanese	0.1220	0.1014	0.0020
Far East Asia	0.1201	0.1007	0.0034
Malaysian	0.1214	0.1012	0.0032
Wine/European	0.1204	0.1009	0.0030
Wine/European - subclade 1	0.1203	0.1011	0.0037
Wine/European - subclade 2	0.1204	0.1011	0.0039
Wine/European - subclade 3	0.1202	0.1010	0.0037
Wine/European - subclade 4	0.1202	0.1011	0.0040
CHNV	0.1212	0.1007	0.0027
Ecuadorean	0.1208	0.1007	0.0035
Far East Russian	0.1209	0.1012	0.0036
North American Oak	0.1206	0.1008	0.0027
Asian Islands	0.1209	0.1011	0.0030
Sake	0.1211	0.1012	0.0029
Asian Fermentation	0.1206	0.1010	0.0024
Alpechin	0.1208	0.1007	0.0035
Brazilian Bioethanol	0.1200	0.1003	0.0029
Mediterranean Oak	0.1204	0.1007	0.0038
French Dairy	0.1192	0.1006	0.0056
African Beer	0.1196	0.1007	0.0037
Mosaic Beer	0.1207	0.1007	0.0035
Mixed Origin	0.1191	0.0994	0.0037
Mexican Agave	0.1200	0.1007	0.0044
M1 - Mosaic Region 1	0.1205	0.1007	0.0033
M2 - Mosaic Region 2	0.1204	0.1007	0.0034
M3 - Mosaic Region 3	0.1205	0.1003	0.0017

B.

S. cerevisiae population	Thermotolerance genes[a]	Genome[b]	p-value[c]
French Guiana Human	0.1259	0.1032	0.0016
Ale Beer	0.1226	0.1015	0.0035
West African Cocoa	0.1258	0.1021	0.0009
African Palm Wine	0.1266	0.1026	0.0019
CHNIII	0.1253	0.1031	0.0018
CHNII	0.1249	0.1020	0.0014
CHNI	0.1271	0.1026	0.0007
Taiwanese	0.1271	0.1031	0.0010
Far East Asia	0.1243	0.1024	0.0014
Malaysian	0.1271	0.1028	0.0008
Wine/European	0.1228	0.1027	0.0056
Wine/European - subclade 1	0.1252	0.1027	0.0022
Wine/European - subclade 2	0.1252	0.1029	0.0018
Wine/European - subclade 3	0.1252	0.1028	0.0014
Wine/European - subclade 4	0.1251	0.1028	0.0025
CHNV	0.1264	0.1028	0.0016
Ecuadorean	0.1266	0.1025	0.0003
Far East Russian	0.1278	0.1028	0.0009
North American Oak	0.1265	0.1025	0.0009
Asian Islands	0.1266	0.1027	0.0008
Sake	0.1278	0.1028	0.0007
Asian Fermentation	0.1261	0.1026	0.0007
Alpechin	0.1259	0.1026	0.0012
Brazilian Bioethanol	0.1251	0.1021	0.0009
Mediterranean Oak	0.1254	0.1025	0.0014
French Dairy	0.1236	0.1024	0.0015
African Beer	0.1243	0.1026	0.0017
Mosaic Beer	0.1257	0.1026	0.0006
Mixed Origin	0.1239	0.1013	0.0017
Mexican Agave	0.1249	0.1023	0.0016
M1 - Mosaic Region 1	0.1256	0.1026	0.0007
M2 - Mosaic Region 2	0.1255	0.1025	0.0023
M3 - Mosaic Region 3	0.1259	0.1020	0.0011

Supplementary Table 3. Divergence between *S. cerevisiae* and *S. paradoxus* at thermotolerance genes. A. [a]Median absolute divergence (D_{xy}) between the indicated population of *S. cerevisiae* from (Peter et al., 2018) and the Wine/European *S. paradoxus* population from (Bergström et al., 2014) in thermotolerance genes. [b]Genomic median D_{xy} for each population pair as in *a.* [c]Empirical significance from a resampling test. **B.** Data are as in **A**, except that each comparison is between the

indicated population of *S. cerevisiae* from (Peter et al., 2018) and the North American *S. paradoxus* subpopulation B from (Durand et al., 2019).

Population	# Isolates	Thermotolerance genes[a]	Genome[b]	*p*-value[c]
Wine/European	362	0.038	0.026	0.0526
Mosaic Region 3	113	0.020	0.019	0.200
Mixed Origin	72	0.076	0.061	0.0493*
Sake	47	0.265	0.258	0.280
Brazilian Bioethanol	35	0.131	0.109	0.0253*

Supplementary Table 4. Thermotolerance loci are enriched for allele-sharing within *S. cerevisiae* populations. Each row reports results from analyses of the allele-sharing statistic genotype homozygosity, G1, for the indicated population of *S. cerevisiae* from (Peter et al., 2018). [a]Median G1 in thermotolerance genes. [b]Genomic median G1. [c]Empirical significance from a resampling test. *, $p < 0.05$.

A

	Genome	Thermotolerance loci	*p*-value
S. paradoxus[a]	0.197	0.138	0.8315
S. uvarum[b]	0.073	0.089	0.6912

B

	Genome	Thermotolerance loci	*p*-value
S. paradoxus[a]	-0.011	0.138	0.2086
S. uvarum[b]	0.001	-0.029	0.4362

Supplementary Table 5. No consistent *cis*-regulatory divergence between *S. cerevisiae* and other species at thermotolerance loci. A. [a]Median *cis*-regulatory divergence in thermotolerance genes between alleles of the indicated species in their F1 hybrid. [b]Data are as in [a] except that the value reports an average across random sets of genes from the genome. [c]Results of a one-sided resampling test for an elevated magnitude of *cis*-regulatory divergence in thermotolerance genes. **B.** Data are as in **A** except that the test was for directional *cis*-regulatory divergence across thermotolerance genes.

Species background	Strain background	Swap (if applicable)[a]	Swap boundaries[a]	Source	Strain Name
S. paradoxus x *S. cerevisiae*	Z1 x DBVPG1373	N/A	N/A	Weiss et al., 2018	CW27
S. paradoxus	Z1	N/A	N/A	Weiss et al., 2018	CW62
S. cerevisiae	DBVPG1373	N/A	N/A	Weiss et al., 2018	CW68
S. cerevisiae	DBVPG1373	ESP1 full swap	Chr7:687097-682525	Weiss et al., 2018	CW98
S. cerevisiae	DBVPG1373	ESP1 promoter swap	Chr7:687463-687842	This study	CW339
S. cerevisiae	DBVPG1373	ESP1 coding swap	Chr7:682525-687419 (heterozygous at position 687419)	This study	CW412
S. cerevisiae	DBVPG1373	ESP1 coding swap	Chr7:682525-687419	This study	CW413
S. cerevisiae	DBVPG1373	ESP1 coding swap	Chr7:682525-687419	This study	CW414

Supplementary Table 6. Strains used in this study. [a]For strains harboring *ESP1* from *S. paradoxus* swapped into the *S. cerevisiae* background at the endogenous location,

listed is the mode of transgenesis (where full indicates both the promoter and coding regions) and boundaries of swapped genetic material. N/A, not applicable.

Strain	Temperature	Minimum[a]	Maximum[a]	Median[a]	Total[b]	Images[c]
Sp, WT	28°C	0	2	0	5	21
	39°C	0	6	2	42	21
Sc, WT	28°C	0	4	2	37	19
	39°C	0	3	1	15	16
ESP1 full swap	28°C	0	6	2	32	19
	39°C	0	2	0	13	27

Supplementary Table 7. Details of microscopy experiments analyzing cell division in wild-type and *ESP1* transgenic yeast. Each row reports the details, from Figure 4 of the main text, of analyses of microscopy of the indicated strain cultured at the indicated temperature. [a]Minimum, maximum, and median dyad count per image. [b]Total number of dyads counted across images. [c]Total number of images scored.

Chapter 3

Barcoded reciprocal hemizygosity analysis via sequencing illuminates the complex genetic basis of yeast thermotolerance

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

Abrams, M.B., Chuong, J.N., AlZaben, F., Dubin, C.A., Skerker, J.M., and Brem, R.B. (2021) Barcoded reciprocal hemizygosity analysis via sequencing illuminates the complex genetic basis of yeast thermotolerance. G3 Genes|Genomes|Genetics.

ABSTRACT

Decades of successes in statistical genetics have revealed the molecular underpinnings of traits as they vary across individuals of a given species. But standard methods in the field can't be applied to divergences between reproductively isolated taxa. Genome-wide reciprocal hemizygosity mapping (RH-seq), a mutagenesis screen in an inter-species hybrid background, holds promise as a method to accelerate the progress of interspecies genetics research. Here we describe an improvement to RH-seq in which mutants harbor barcodes for cheap and straightforward sequencing after selection in a condition of interest. As a proof of concept for the new tool, we carried out genetic dissection of the difference in thermotolerance between two reproductively isolated budding yeast species. Experimental screening identified dozens of candidate loci at which variation between the species contributed to the thermotolerance trait. Hits were enriched for mitosis genes and other housekeeping factors, and among them were multiple loci with robust sequence signatures of positive selection. Together, these results shed new light on the mechanisms by which evolution solved the problems of cell survival and division at high temperature in the yeast clade, and they illustrate the power of the barcoded RH-seq approach.

INTRODUCTION

Understanding how and why organisms from the wild exhibit different traits is a central goal of modern genetics. Linkage and association mapping have driven decades of success in dissecting trait variation across individuals of a given species (Ott et al., 2015; Tam et al., 2019). But since these methods can't be applied to reproductively isolated taxa, progress in the field of interspecies genetics has lagged behind. However, newer statistical-genetic methods appropriate to comparisons between species have been proposed in the recent literature (Weiss and Brem, 2019), which hold promise for elucidating the genetics of ancient traits. For most such methods, limitations accruing from throughput and/or coverage issues remain to be refined.

The budding yeast *Saccharomyces cerevisiae* grows better at high temperature than any other species in its clade (Gonçalves et al., 2011; Hittinger, 2013; Salvadó et al., 2011; Sweeney et al., 2004; Weiss et al., 2018a), in keeping with its likely ecological

origin in hot, East Asian locales (Peter et al., 2018). This derived and putatively adaptive trait serves as a model for the genetic study of deep evolutionary divergences. Thermosensitivity, the ancestral phenotype in the clade, is borne out in *S. paradoxus*, a close sister species to *S. cerevisiae*, making the former a useful point of comparison. Our group previously used this system as a testbed to develop RH-seq (Weiss et al., 2018a), a genomic version of the reciprocal hemizygosity test (Stern, 2014) that is well-suited to the mapping of natural trait variation between sister species. This technique starts with the generation of large numbers of random transposon mutant clones of a viable but sterile interspecies hybrid. In a given clone, loss of function from a transposon insertion in one species' allele of a gene reveals the function of the uncovered allele from the other species. These hemizygotes are competed en masse in a condition of interest; the abundance of each hemizygote in turn in the selected pool is quantified by bulk sequencing, and used in a test for allelic impact on the focal trait. In previous work, we identified eight genes through this approach at which species divergence contributed to thermotolerance (Weiss et al., 2018a).

Against a backdrop of successful biological and evolutionary inference from our yeast RH-seq pilot (Abrams et al., 2021a; Weiss et al., 2018a), we noted that the combination of *S. cerevisiae* alleles of all eight genes mapped to thermotolerance recapitulated only <20% of the difference between the species (AlZaben et al., 2021). Thus, many of the determinants of yeast thermotolerance likely remain undetected. If so, boosting the replication and throughput of genetic mapping, to enable higher statistical power, could help meet the challenge. In our initial implementation of RH-seq, we had quantified the abundance of hemizygotes in a sample by sequencing across the transposon junction with the genome, using one universal primer that recognized the transposon and another recognizing a ligated adapter at DNA fragment ends (Weiss et al., 2018a). This protocol, though rigorous, is labor-intensive and expensive, limiting the potential for throughput and coverage. A higher-throughput alternative starts with the tagging of transposon sequences by random short DNA barcodes (Wetmore et al., 2015). After mutagenesis of a genotype of interest by these barcoded transposons, and then selection of the mutants in bulk in a challenging condition, mutant abundance can be quantified from sequencing of DNA straight from the pool with a simple PCR. We set out to adapt this barcoding strategy to enable highly replicated RH-seq, with application to yeast thermotolerance as a test case to achieve a deeper exploration of the complex genetics of the trait.

MATERIALS AND METHODS

Construction of a randomly barcoded piggyBac transposase pool

For barcoded RH-seq, we constructed a pool of plasmids, each harboring the piggyBac transposase and a randomly barcoded copy of the piggyBac transposon, via Golden Gate cloning of random 20bp barcodes flanked by universal priming sites into a plasmid backbone containing the piggyBac machinery, modified from pJR487 (Weiss et al., 2018a) as follows (Figure S1).

Preparation of the backbone vector

To allow the use of BbsI as the Type IIS restriction enzyme for Golden Gate cloning of barcodes into pJR487 (see below), we first removed all three BbsI cut sites from pJR487 by introducing silent mutations that disrupted the restriction enzyme's recognition pattern. The resulting plasmid was called pCW328. We next modified pCW328 to make a Golden-Gate-ready vector, with the final identifier pJC31, by replacing transposon nucleotides with those of a stuffer at a location 70 nucleotides from the end of the right arm of the transposon; see Supplementary Note and Figure S2 for a description of this choice. The stuffer contained two BbsI cut sites with custom Type IIS overhang sequences from (Lee et al., 2015), and a NotI cut site in between the two BbsI cut sites. All cloning steps were carried out by GenScript, Inc.

Preparation of barcode oligonucleotides

To make barcodes, we acquired an oligonucleotide pool from IDT that contained random 20 bp sequences (from hand-mixed random nucleotides) flanked by universal priming regions, U1 and U2 (Wetmore et al. 2015, Coradetti et al. 2018). These custom oligos were produced and PAGE purified by IDT. Additionally, we designed forward (FW_BbsI_JC) and reverse (REV_BbsI_JC) primers which each contained a BbsI cut site, BbsI overhang sequences complementary to the backbone vector, and either universal priming sequence (Table S2) (Coradetti et al. 2018). We set up 50 µL amplification PCR reactions with 1 µL of random 20 bp barcodes as template, from a 2.5 µM stock, and 0.25 µL of each of the forward and reverse primers from a 100 µM stock. Amplification used Phusion High Fidelity polymerase (NEB) and the following cycling protocol: 98°C for 30 seconds, (98°C for 10 seconds, 58°C for 30 seconds, 72°C for 60 seconds) × 6, 72°C for five minutes. PCR products were purified (Zymo DNA Clean & Concentrator kit) and then combined. This yielded the final donor barcodes: random 20bp barcodes flanked by universal priming regions, with BbsI cut sites at the extreme edges.

Cloning barcodes into plasmids

To clone barcodes into pJC31, we proceeded in two barcoding reactions.

The first reaction contained 2:1 molar ratio of vector to barcodes (4 µg of pJC31 and 128 ng of donor barcodes), 5 µL of 10X T4 Ligase Buffer (ThermoFisher), 2.5 µL of T4 Ligase (ThermoFisher), 2.5 µL FastDigest BpiI (ThermoFisher), and sterile water up to 50 µL. The cycling program was: 37°C for five minutes, (37°C for two minutes, 16°C for five minutes) x 25, 65°C for 10 minutes. Then a mixture containing 5 µL 10X FastDigest Buffer (ThermoFisher), 3.13 µL BSA 2 mg/mL (NEB), 12.5 µL FastDigest NotI (ThermoFisher), and 12.5 µL FastDigest BpiI (ThermoFisher) was spiked into the reaction and incubated at 37°C for 16 hours to digest unbarcoded backbone vectors. Ten of these reactions were combined, purified, and eluted in H_2O (Zymo DNA Clean & Concentrator). To spot-check this cloning, 5 µL of this product was transformed into 25 µL of *E. coli* 10beta electrocompetent cells (NEB). Sanger sequences across the

barcode regions of 20 individually miniprepped *E. coli* colonies showed 95% barcoding efficiency.

The second reaction contained 2:1 molar ratio of vector to donor barcodes (4µg of pJC31 and 128 ng of donor barcodes), 5 µL of 10X T4 Buffer (ThermoFisher), 2.5 µL T4 Ligase (ThermoFisher), 2.5 µL BpiI (ThermoFisher), and sterile water up to 50 µL. The cycling program was: 37°C for five minutes, (37°C for two minutes, 16°C for five minutes) x 25, 65°C for 10 minutes. Then a mixture containing 2.5 µL 10X FastDigest Buffer (ThermoFisher), 2.5 µL G Buffer, (ThermoFisher), 3.13 µL BSA 2 mg/mL (NEB), 12.5 µL FastDigest NotI (ThermoFisher), and 12.5 µL BpiI (ThermoFisher) was spiked in the reaction and incubated at 37°C for 16 hours to digest remaining unbarcoded backbone vectors. Six of these reactions were combined, purified, and eluted in H_2O (Zymo DNA Clean & Concentrator). Then every 5 µL of cleaned eluted product was redigested with 5µL of NotI-HF (NEB), 5 µL 10X CutSmart buffer (NEB), and 35 µL H_2O at 37°C for 16 hours then 80°C for 20 minutes. The reactions were purified again (Zymo DNA Clean & Concentrator) and pooled. Spot checks of this cloning reaction proceeded as above, and Sanger sequences across the barcode regions of 20 individually miniprepped *E. coli* colonies showed 95% barcoding efficiency.

Purified plasmids from the two reactions were combined in a master tube of DNA before transforming into electrocompetent *E. coli* cells (NEB) to generate the final barcoded piggyBac pool (final identifier P58). Each electroporation cuvette (BTX) contained 25 µL of 10beta electrocompetent cells (NEB) and 5 µL of cleaned master tube DNA from the previous golden gate barcoding step. We performed 21 electroporation reactions in total using the Bio-Rad GenePulser Xcell machine set to 2.0 kV, 200 Ohms, 25 µF. After electroporation, each culture was recovered in provided outgrowth media (NEB) by shaking at 37°C at 250 rpm for 1.5 hours. After recovery, all independent 21 electroporation reactions were combined.

The combined recovered transformation *E. coli* culture was used to inoculate two 1L fresh LB cultures containing carbenicillin at 100 µg/mL to select for *E. coli* cells containing barcoded piggyBac plasmids. Each culture was incubated for 15.5 hours at 37°C, 250 rpm (overnight) to expand the barcoded piggyBac *E. coli* pool. Then the two cultures were combined yielding the final barcoded transposon plasmid pool, P58. This was aliquoted into 1 mL volumes with 15% glycerol and stored at -80°C.

Sequencing verification of barcoded piggyBac pool plasmid DNA for barcode diversity
To verify barcode diversity in the barcoded piggyBac plasmid pool (P58), we sequenced barcodes as follows. One frozen aliquot of P58 was inoculated into 1.25 L of LB containing carbenicillin 100 µg/mL and grown for 16 hours 37°C, 250 rpm or until it reached an OD_{600} of 2.1. This culture was gigaprepped on using a column kit (Invitrogen) to generate 5 mg of plasmid. We used this as input into a PCR with primers (Table S2) annealing to the universal priming regions flanking the barcode. These primers were dual-indexed, although in this work we only carried out sequencing of the resulting amplicon from one end (see below), such that only one index was used. The generic form of the forward primer was

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT CT(N1-4)xxxxxxGTCGACCTGCAGCGTACG, where the N1-4 represent variable amounts of random bases from 1-4 to help samples cluster on the Illumina lane and the (x6) represent a unique 6-bp index sequence for multiplexing samples. The generic reverse primer was CAAGCAGAAGACGGCATACGAGATxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTT CCGATCTGATGTCCACGAGGTCTCT . Four PCR reactions used 50 ng of prepped P58 plasmid template each. Amplification used Q5 High Fidelity Polymerase (NEB) and a cycling program 98°C for four minutes, (98°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds) x 25, 72°C for five minutes. Each PCR product was purified on a column (Zymo DNA Clean & Concentrator-5 Kit) and eluted in 10 µL prewarmed 65°C provided elution buffer (Zymo). Six µL of each were then combined and sequenced off the U2 region via Illumina amplicon sequencing, on one lane of HiSeq4000 SR50 at the Genomics Sequencing Laboratory at UC Berkeley. Reads sequenced per library are reported in Table S3. Sequencing of the *E. coli* vector pool p58 revealed 27,538,142 barcodes with an estimated sequencing error rate of 1.38% analyzed as described (Coradetti et al., 2018).

Yeast hemizygote pool construction via barcoded transposon mutagenesis

We constructed our yeast hemizygote pool essentially as described (Weiss et al., 2018a) but with modifications as follows.

To prepare plasmid DNA for mutagenesis, one frozen aliquot of P58 was inoculated into 1.25L of LB containing carbenicillin 100 µg ml^{-1} and grown for 16 hours at 37°C, 250 rpm or until it reached an OD_{600}/mL of 2.1. This culture was gigaprepped on using a column kit (Invitrogen) to generate 5 mg of plasmid.

Next, we transformed yeast in several, smaller subpools which we combined to form a final pool as follows. We carried out mutagenesis of CW27, an F1 hybrid from the mating of *S. cerevisiae* DBVPG1373 with *S. paradoxus* Z1 (Weiss et al., 2018a) across the first two days. The first day, we generated one subpool in a single 50 mL culture and one subpool in five 50 mL cultures at OD_{600}/mL ~0.9 (~45 OD_{600} units of cells each). The second day, we generated two subpools in five 50 mL cultures each at OD_{600}/mL ~0.9 (~45 OD_{600} units of cells).

To generate subpools consisting of a single 50 mL culture, one colony of CW27 was inoculated into 5 mL of YPD and incubated at 28°C 200 rpm. 24 hours later, the OD_{600}/mL of the overnight culture was 3.86. It was backdiluted to an OD_{600}/mL of 0.1 in 50 mL of YPD in a 250 mL Erlenmeyer flask and grown with shaking at 28°C, 200 rpm for 5.5 hours. After 5.5 hours, it had reached OD_{600}/mL ~0.9 and cells were at mid-log phase. This 50 mL culture was gently pelleted at 1000xg for three minutes. The pellet was washed with 25 mL sterile water and then 5 mL of 0.1 M lithium acetate (Sigma) mixed with 1X Tris-EDTA buffer (10 mM Tris-HCl and 1.0 mM EDTA); after spin-down, to the tube was added a solution of 0.269 mg of P58 mixed 5:1 by volume with salmon sperm DNA (Invitrogen), followed by 3 mL of 39.52% polyethylene glycol, 0.12 M lithium

acetate and 1.2X Tris-EDTA buffer (12 mM Tris-HCl and 1.2 mM EDTA). The tube was rested for 10 minutes at room temperature, then heat-shocked in a water bath at 37°C for 26 minutes. The tube was gently spun at 1000g for three minutes after which supernatant was removed. We transferred the cells to a flask and added YPD to attain an OD_{600}/mL of ~0.35–4 in ~70 mL. Each such culture was recovered by shaking at 28°C and 200 rpm for two hours. G418 (Geneticin; Gibco) was added to each at a concentration of 300 µg/mL to select for those cells that 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}/mL of ~2.5. We transferred cells from this culture, and YPD + G418 (300 µg/mL), to new 250 mL flasks at the volumes required to attain an OD_{600}/mL of 0.2 in 50 mL each. We cultured each flask with 200 rpm. shaking at 28°C overnight until each reached an OD_{600}/mL of 3.43. To cure transformants of the P58 URA3+ plasmid, we spun down 10% of this master culture, and resuspended in water with the volume required to attain a cell density of 1.85 OD_{600}/mL. Four mL of this resuspension were plated (1 mL per 24.1 cm x 24.1 cm plate) onto plates containing complete synthetic media with 5-fluorooritic acid (0.2% dropout amino acid mix without uracil or yeast nitrogen base (US Biological), 0.005% uracil (Sigma), 2% D-glucose (Sigma), 0.67% yeast nitrogen base without amino acids (Difco), 0.075% 5-fluorooritic acid (Zymo Research)). After incubation at 28°C to enable colony growth, colonies were scraped off all four plates and combined into water at the volume required to attain 44 OD_{600}/mL, yielding the transposon mutant hemizygote subpool. This was aliquoted into 1 mL volumes with 10% dimethylsulfoxide and frozen at −80°C.

To generate subpools consisting of five 50 mL cultures, one colony of CW27 was inoculated to 100 mL of YPD in a 250 mL Erlenmeyer flask and incubated shaking at 28°C, 200 rpm. Twenty-four hours later, the OD_{600}/mL of the overnight culture was OD_{600}/mL 3.89. The overnight culture was backdiluted to OD_{600}/mL 0.1 in 250 mL of YPD and incubated for 5.5 hours at 28°C, 200 rpm. After 5.5 hours, the OD_{600}/mL reached 0.9 and cells were split into five 50 mL conical tubes, and subjected each to heat shock as above. We then transferred all cells from this post-transformation culture to one 1L flask and added fresh YPD to attain OD_{600}/mL 0.4 in ~750 mL YPD. The transformed culture was recovered by shaking at 28°C, 200rpm, for two hours. G418 (300mg/ul) was added to select for the transposed cells. The culture continued shaking for 48 hours or until the OD_{600}/mL reached 2.1. This culture was then backdiluted to create a new culture at OD_{600}/mL 0.2 in 500 mL of YPD with 300mg/µL G418 shaking for 24 hours at 28°C, 200 rpm until it reached OD_{600}/mL ~3.4 The curing, scraping, and freezing steps were the same as above.

To combine the four subpools to yield the final 160X hemizygote pool (final identifier P75), three 1 mL aliquots of each subpool were thawed on ice for one hour. They were transferred to each of four 1L flasks with 500 mL YPD to OD_{600}/mL 0.2, cultured at 28°C, 200 rpm for 17 hours upon which the OD_{600}/mL was 3.5-4. They were gently pelleted, combined, and resuspended in two ways to reach OD_{600}/mL of 44: YPD with 15% glycerol and YPD with 7% DMSO, aliquoted to 1 mL volumes, and frozen at -80°C.

Tn-seq mapping of yeast hemizygote pool

Tn-seq library preparation

To associate barcoded transposon insertions to genomic location in the hemizygote pool, which we refer to as Tn-seq, we first sequenced barcoded transposon insertions according to the methods of (Weiss et al., 2018a) as follows. Each 44 OD_{600}/mL aliquot of each subpool or final pool 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 Research) prewarmed to 65°C, and its concentration was quantified using a Qubit 4.0 fluorometer. Illumina transposon sequencing (Tn-seq) library construction was as described previously. Briefly, gDNA was sonicated and ligated with common adapters, and for each fragment deriving from a barcoded transposon insertion in the genome, a sequence containing a barcode, a portion of the transposon, and a portion of its genomic context (the barcoded transposon–genome junction) was amplified using one primer homologous to the U1 region immediately upstream of barcode and another primer homologous to a region in the adapter. See Table S2 for the transposon-specific primer ("forward primer"), where Ns represent random nucleotides, and the indexed adapter-specific primer ("reverse primer"). Amplification used Jumpstart polymerase (Sigma) and the following cycling protocol: 94°C for two minutes, (94°C for 30 seconds, 65°C for 20 seconds, 72°C for 30 seconds) × 25, 72°C for 10 minutes. Sequencing of paired-end reads of 150 bp was done over two lanes on a HiSeq4000 at Novogene Corporation (Sacramento, CA) and one lane on a NovaSeq SP at the Genomics Sequencing Laboratory at UC Berkeley (Berkeley, CA). Reads sequenced per library are reported in Table S4.

Tn-seq data analysis

Tn-seq data of the hemizygote pool was analyzed, to infer transposon insertions on the basis of barcodes detected in reads as junctions with genomic sequence, essentially as described (Coradetti et al., 2018) (https://github.com/stcoradetti/RBseq/tree/master/Old_Versions/1.1.4), with the following modifications. For each barcode, instead of scanning positions for the end of the insertion from a sequence specified by a model file, we searched for the final 22 base pairs of the right arm of the piggyBac transposon allowing for two mismatches. For annotation, we converted the annotation file from https://github.com/weiss19/rh-seq for the *S.* cerevisiae D1373 x *S. paradoxus* Z1 hybrid to a compliant GFF3 file using Another GFF Analysis Toolkit (AGAT) - Version: v0.4.0 (https://github.com/NBISweden/AGAT). Then, we used a custom Jupyter notebook to annotate the file generated by the RBseq mapping software.

Quality control for Tn-seq, to eliminate barcodes whose junction genomic sequence mapped to multiple insertion locations in the hybrid genome, and to minimize the proportion of sequencing errors included in final tallies, was as described (Coradetti et al., 2018). Briefly, we eliminated from further consideration any case where a barcode observed in Tn-seq sequencing data differed from another, much more abundant, barcode by a single base (a total of 2,024,812 off-by-one barcodes in 2,888,129 reads).

We also filtered out off-by-two barcodes (280,949 barcodes in total). Separately, we eliminated barcodes that were detected in sequencing data as a junction with more than one genomic context, suggesting the respective transposon had inserted into multiple locations in one or many clones (98,669 barcodes where this inference was based on multiple strong mapping matches, and an additional 46,583 barcodes where this inference was ambiguous, with one strong mapping match with reads outnumbered by those assigned to weaker mapping matches). The final filtered barcode set comprised 548,129 uniquely barcoded and mapped inferred transposon insertions in the P75 hemizygote pool, at an average read depth of 308.6 reads, and a median read depth of 47 reads; 166,834 of these insertions were mapped as genic. The annotation script, GFF3 file, and modified mapping script are available at https://github.com/melanieabrams-pub/RH-seq_with_barcoding.

Competition cultures

For the thermotolerance competition at 37°C, one aliquot of the yeast hemizygote pool was thawed and inoculated into 150 mL of YPD in a 250 mL unbaffled Erlenmeyer flask and grown for six hours at 28°C, 200 rpm. This pre-culture (T_0, at OD_{600}/mL of 1.22) was backdiluted into 12 10 mL competition cultures at 200 rpm at each of 28°C and 37°C, with a starting OD_{600}/mL of 0.02 or 0.05 in at 28°C and 37°C respectively. These competition cultures were maintained within logarithmic growth through back-dilutions into fresh tubes of 10 mL of YPD at the same optical density as the starting culture, for a total of 10-15 generations. Dilutions for the 28°C competition cultures were performed after 8.5, 18.5, and 25.5 hours after the T_0 timepoint, and dilutions for the 37°C competition cultures were performed after 8.5, 18.5, 25.5 hours and 32.5 hours after the T_0 timepoint. The entire cell culture was harvested from each of these biological replicate tubes for sequencing as biological replicates.

For thermotolerance competition at 36°C, competition cultures were grown as above with the following differences. The high temperature was 36°C, instead of 37°C. The pre-culture (T_0, at OD_{600}/mL of 0.693 after 5.5 hours at 28°C, 200 rpm) was backdiluted to a starting OD_{600}/mL of 0.02 for competition cultures at 36°C. Dilutions for both the 28°C and 36°C competition cultures were performed after 8.5, 18.5 and 25.25 hours after the T_0 timepoint. Eleven instead of 12 replicates were carried out at 28°C.

Barcode quantification from competition cultures

Bar-seq library preparation

To determine the abundance of barcoded transposon mutant hemizygote clones after selection, we sequenced barcodes insertions as follows. Each cell pellet from a selection sample was thawed on ice, and its genomic DNA (gDNA) was harvested with the Zymo QuickDNA Kit (Zymo#D6005). gDNA was resuspending in DNA elution buffer (Zymo Research) prewarmed to 65°C, and its concentration was quantified using a Qubit 4.0 fluorometer. The barcode insertion was amplified as above (see *Sequencing verification of barcoded piggyBac pool plasmid DNA for barcode diversity*). Each PCR

product was purified on a column (Zymo DNA Clean & Concentrator) and eluted in 10 µL prewarmed 65°C provided elution buffer (Zymo). Six µL of each were then combined and sequenced off the U2 region by Illumina sequencing on one lane of HiSeq4000 SR50 at the QB3 Genomics Sequencing Laboratory at UC Berkeley.

Bar-seq data analysis

Bar-seq mapping and quantification were as described (Coradetti et al., 2018) (https://github.com/stcoradetti/RBseq/tree/master/Old_Versions/1.1.4), wherein only barcodes that passed quality control in Tn-seq (see *Tn-seq data analysis* above) were analyzed for quantitative measures of abundance via Bar-seq. Thus we did not use in our screen any barcode that was detected in Bar-seq sequence data but not Tn-seq data (the product of *e.g.* sequencing errors in Bar-seq, or a failure to observe in Tn-seq a barcode associated with a bona fide transposon insertion that could be detected in Bar-seq). A total of 301,349 barcodes conformed to these criteria from across all replicates of Bar-seq in competitions for the dissection of determinants of growth at 37°C relative to 28°C, with an average read depth of 305.3 reads and a median of 12 reads; 89,772 of these Bar-seq detected barcodes corresponded to inferred transposon insertions in genes and were analyzed as input to the reciprocal hemizygosity testing pipeline described below. In a given replicate competition culture we detected a median 1×10^5 barcodes. The latter represented a fifth of the size of the total pool of hemizygotes detectable after quality control by Tn-seq (5.5×10^5; see *Tn-seq data analysis* above). Thus, the extent of bottlenecking in any given competition experiment was modest, with diversity retained at the order of magnitude of the mutant pool size.

Competitions for the dissection of growth at 36°C relative to 28°C used the same procedures as above, mapping a total of 230,469 barcodes, 68,523 of which corresponded to inserts in genes and were analyzed as input to the reciprocal hemizygosity testing pipeline described below. In a given replicate competition culture, we detected a median 5×10^4 barcodes.

Reciprocal hemizygosity testing

The tabulated counts of abundance from Bar-seq for each barcode in each replicate were used as input into reciprocal hemizygosity tests essentially as in (Abrams et al., 2021a), with slight changes as follows. We had in hand each barcode which had been sequenced as a junction with a unique genomic location in the Tn-seq step and had passed quality control there (see *Tn-seq data analysis* above), and which was now detected in competition cultures. We interpreted each such barcode as reporting a hemizygote clone bearing a transposon insertion at the respective position of the respective species' allele (*S. cerevisiae* or *S. paradoxus*), with the other species' allele retained as wild-type at that locus. In what follows, we refer to each such barcode as reporting an inferred hemizygote clone, with respect to its growth behavior in competition cultures. As in ((Abrams et al., 2021a)), for a given biological replicate we normalized the abundances attributed to each inferred hemizygote genotype to the total number of sequencing reads in the respective sample, and we eliminated from further

analysis insertions which had been annotated as intergenic, or as corresponding to the plasmid used to generate this library. For reciprocal hemizygosity tests, we excluded from consideration any gene with fewer than three inferred hemizygote genotypes per allele. Of the retained genes, for each inferred hemizygote genotype, we tabulated the quantity $a_{experimental,i}$, the sequencing-based abundance measured after the competition culture in biological replicate *i* of growth at the experimental temperature (36 or 37°C), and, separately, we calculated $a_{control,i}$, the analogous quantity from growth at the control temperature (28°C), for *i* = [1,12]. We then took the mean of the latter and used it to tabulate the temperature effect on the inferred hemizygote genotype in replicate *i*, t_i = $\log_2(a_{experimental,i}/a_{control,mean})$. As in ((Abrams et al., 2021a)), we eliminated an inferred hemizygote genotype if the coefficient of variation of this quantity exceeded 2.0, or there were fewer than 1.1 normalized reads. With the data for the remaining inferred hemizygote genotypes, for a given gene, we compiled the vector of the *t* measurements across replicates and all inferred hemizygote genotypes with each species' allele of the hybrid disrupted in turn, and discarded genes where the coefficient of variation of the *t* measurements across hemizygote inserts for one or both alleles exceeded 10. For the remainder, we used the Mann-Whitney test to compare these two vectors, with Benjamini-Hochberg correction for multiple testing. For a given gene, we calculated the effect size as the difference between two values: the $\log_2$(abundance at the experimental temperature/abundance at 28°C) of the average inferred hemizygote genotype representing a transposon insertion in the *S. cerevisiae* allele, and the analogous quantity among inferred hemizygote genotypes representing insertions in the *S. paradoxus* allele of the gene. Scripts for this modified RH-seq analysis pipeline are available at https://github.com/melanieabrams-pub/RH-seq_with_barcoding. We earmarked top candidate genes for factors contributing to the thermotolerance of *S. cerevisiae* as those with corrected Mann-Whitney $p < 0.05$ in the reciprocal hemizygosity test, and an effect size < -0.5, *i.e.* disrupting the *S. cerevisiae* allele was associated with a strong defect in thermotolerance relative to disruption of the *S. paradoxus* allele; we refer to this gene set as our top barcoded RH-seq hit gene list.

Analysis of inferred interactions between top hit genes from barcoded RH-seq

For the circos plot reporting inferred interactions between top hit genes from barcoded RH-seq, we used the STRING database (Szklarczyk et al., 2021), accessed September 30, 2021, which incorporates experimental/biochemical data from DIP, BioGRID, HPRD, IntAct, MINT, and PDB, and curated data from Biocarta, BioCyc, Gene Ontology, KEGG, and Reactome. Widths of edges between nodes in the circos plot represent STRING confidence scores, each the probability of a true positive interaction between a given two genes (Szklarczyk et al., 2021).

To test the encoded proteins of top barcoded RH-seq hit genes for enrichment of physical interactions with each other, we used curated known interactions from BioGRID (Oughtred et al., 2021) as housed in the Saccharomyces Gene Database, downloaded February 19, 2021. We tabulated the number of physical interactions between the proteins encoded by RH-seq hit genes, and we divided that by the total number of interactions involving one RH-seq hit gene and any other gene in the

genome; call this ratio r_{true}. Then, we drew a random sample of genes from the genome, as described above for GO term resampling. We tabulated, in this random gene set, the number of physical interactions between genes in that sample, and we divided that by the total number of interactions involving one gene in the random sample and any other gene in the genome, to yield $r_{resample}$. We repeated this procedure 10,000 times, and we used the proportion of resampled groups where $r_{resample}$ was greater than or equal to r_{true} as a one-sided *p* value assessing the significance of enrichment of interactions between our genes of interest.

Gene Ontology analyses of top hit genes from barcoded RH-seq

To test top barcoded RH-seq hit genes for enrichment for overrepresentation of a particular Gene Ontology (GO) term, we mapped each gene to its Gene Ontology groups based on data from geneontology.com (Ashburner et al., 2000). We filtered out GO terms with fewer than five or with more than 200 gene members. We also filtered out GO terms with identical membership in the genome. We took the subset of the remaining GO terms with at least one member among our top barcoded RH-seq hit genes. Then, we randomly sampled genes from the genome, ensuring the same proportion of essential genes as in our set of top barcoded RH-seq hit genes based on the essentiality annotations of (Winzeler et al., 1999). We tabulated whether our random sample had greater or fewer genes with that GO term than our candidate set. We repeated this procedure 10,000 times and used the proportion of these resampled groups that had more genes in the given GO term as the initial *p* value assessing the significance of the enrichment of that GO term. Then, we applied Benjamini-Hochberg correction for multiple hypothesis testing to generate final, adjusted *p*-values for the enrichment of the given GO term among top barcoded RH-seq hit genes.

To test Biological Process ontologies for enrichment for large magnitudes of the effect of allelic variation on thermotolerance, we used the latter as tabulated in **Reciprocal hemizygosity testing** above. We filtered GO terms as above, and then excluded all genes absent in our barcoded RH seq analysis, which would have no associated quantity for the effect of allelic variation to resample. For each retained term in turn, we first tabulated the median absolute value of the effect size of the gene members for which we had data, e_{true}. Then, we tabulated the analogous quantity for a random sample of the same number of genes from the genome, $e_{resample}$, ensuring the same proportion of essential genes as above. We repeated this procedure 100 times, and used the proportion of the resampled groups for which $e_{resample}$ was greater than or equal to e_{true} as an initial *p* value assessing the enrichment of large effects of allelic variation in the genes the term. For all GO terms with an initial *p* value < 0.1, we repeated this procedure 10,000 times to calculate a more precise *p* value. Then, we applied the Benjamini-Hochberg correction for multiple hypothesis testing to generate final, adjusted *p*-values for the enrichment of the given GO term for large effects of allelic variation on thermotolerance.

Molecular evolution analysis of RH-seq hit genes

Branch length PAML analysis with codeML was performed as in (Dubin et al., 2020). Hits were manually inspected for the quality of the alignment, and one, *YAL026C*, was discarded for poor alignment quality leading to an artifactually high branch length. We used the inferred branch lengths as input into a resampling test as in **Gene Ontology analyses of top hit genes from barcoded RH-seq** above, and we performed a one-sided significance test for long branch lengths along the *S. cerevisiae* lineage. Branch-site PAML analysis with codeML was performed as in (Abrams et al., 2021a). Jalview version 2 was used to visualize the percent identity of amino acid sequence alignments (Waterhouse et al., 2009). McDonald-Kreitman analysis statistics were calculated as in (Abrams et al., 2021a). Fisher's exact test was used to compute *p*-values for individual loci, and these were adjusted using the Benjamini-Hochberg correction for multiple hypothesis testing.

RESULTS

Dissecting thermotolerance divergence between species by barcoded transposon mutagenesis

With the goals of boosting RH-seq throughput and power, and achieving new insights into the genetics and evolution of yeast thermotolerance, we set out to generate an RH-seq reagent for yeast incorporating barcoded transposons (Wetmore et al., 2015). For this purpose, we first generated a pool of plasmids, each encoding a barcoded copy of the piggyBac transposon and its transposase (Figure S1A-C). To use these in RH-seq, we revisited our previously characterized model system: a comparison between DVBPG1373, a thermotolerant Dutch soil strain of *S. cerevisiae,* and Z1, an *S. paradoxus* isolate from the UK (Abrams et al., 2021a; AlZaben et al., 2021; Weiss et al., 2018a). The F1 hybrid formed from the mating of these strains exhibits a thermotolerance phenotype intermediate between those of the two species parents, and thus is well-suited to mapping of allelic effects on the trait (Weiss et al., 2018a). We transformed this F1 hybrid with barcoded plasmids, yielding a pool of hemizygote mutants, which we expanded and then banked (Figure S1D). Next, to catalog the genomic locations of transposon insertions, we used the DNA from a culture of the pool in standard conditions as input into a first round of sequencing library construction, whose primers recognized a common site on the transposon and a common DNA adapter ligated to DNA fragment ends ("Tn-seq"; Figure 1A). Sequencing and data analysis, with quality controls to eliminate barcodes that could not be uniquely associated with a single transposon insertion location (see Methods), yielded a catalog of 548,129 barcoded hemizygotes in the pool whose genomic insertion locations were tabulated. At this point we could harness the pool for highly replicated screens, each of which could quantify hemizygote abundance in a condition of interest via relatively cheap and straightforward barcode sequencing ("Bar-seq"; Figure 1B).

Thus, with our barcoded hemizygote pool, we implemented an RH-seq screen to search for genes at which *S. cerevisiae* and *S. paradoxus* alleles drove differences in strain abundance at high temperature. For this, we subjected the pool to growth assays with 12 biological replicate cultures at 37°C, alongside controls at 28°C. We used DNA from

each culture as input into barcode sequencing (Figure 1B). The resulting data revealed a total of 301,349 cases where a barcode, representing a hemizygote clone with a transposon insertion catalogued by Tn-seq (Figure 1A), was detectable in our growth assays. Transposon insertion positions corresponding to these informative barcodes were evenly split between *S. cerevisiae* and *S. paradoxus* alleles of genes throughout the F1 hybrid genome (Figure S3). We took the normalized count of a given barcode in a sequencing data set as a report of the fitness of the respective hemizygote, *i.e.* its relative abundance after growth in the pool in the respective condition. We then used the complete set of such counts as the input into reciprocal hemizygosity tests to compare, for a given gene, the temperature-dependent abundance of strains harboring a disruption in the *S. cerevisiae* allele, relative to that of strains with the *S. paradoxus* allele disrupted. A pipeline for these tests, including filters for coverage and reproducibility and multiple testing correction (see Methods), revealed 83 genes at a 5% false discovery rate. This contrasted with the much smaller set of eight genes at which species' alleles drove differences in high-temperature growth, in our original non-barcoded RH-seq approach (Weiss et al., 2018a), which had involved only three biological replicates. The 10-fold increase in the number of significant hits in our barcoded RH-seq screen reflects the statistical power afforded by our highly-replicated method to detect even quite small effects.

In our barcoded RH-seq screen hits, as a positive control we first examined the set of genes known to contribute to thermotolerance divergence from our earlier study (*AFG2, APC1, CEP3, DYN1, ESP1, MYO1, SCC2*, and *DYN1*) (Weiss et al., 2018a). Several did not meet the experiment-wide statistical thresholds of our barcoded RH-seq pipeline (Figure S4A), suggesting an appreciable false negative rate of the latter overall. However, manual inspection made clear that hemizygosity effects at all gold-standard thermotolerance loci were borne out: in each case, in barcoded RH-seq data, strains with disruptions in the *S. cerevisiae* allele, and a wild-type copy of the *S. paradoxus* allele, had worse thermotolerance than did strains with only the *S. cerevisiae* allele intact (Figure S4A-B), as we had previously reported (Weiss et al., 2018a). Furthermore, the list of gene hits from barcoded RH-seq also included *HFA1* (Figure 2B and Table S6) which was reported and validated separately as a determinant of thermotolerance differences between yeast species (Li et al., 2019). On the strength of these controls, we considered our deep sampling of thermotolerance loci to serve as a useful proof of concept for the barcoded RH-seq method.

Functional-genomic analysis of thermotolerance genes

We next aimed to pursue deeper analyses of the novel gene hits from barcoded RH-seq in our yeast thermotolerance application. We considered that a focus on the strongest and most evolutionarily relevant sources of mapping signal would likely yield the most informative results. As such, in light of our interest in explaining the exceptional thermotolerance of purebred *S. cerevisiae*, we earmarked the 44 genes from our larger candidate set at which the *S. cerevisiae* allele boosted the trait most dramatically relative to that of *S. paradoxus* (Figure 2 and Table S6). In what follows, we refer to these genes as our top RH-seq hits, and we analyze them as our highest-confidence

predictions for factors that nature would have used in evolving the *S. cerevisiae* phenotype.

We sought to use our mapped loci to explore potential functional mechanisms underlying the thermotolerance trait. We hypothesized that *S. cerevisiae* thermotolerance genes could participate in an interacting network, jointly shoring up particular aspects of cell machinery that were critical for growth at high temperature (AlZaben et al., 2021). Consistent with this notion, the STRING database, which collates experimentally detected protein-protein interactions, genetic interactions, and pathway membership (Szklarczyk et al., 2021), inferred multiple interactions among our top genes from barcoded RH-seq, with salient signal involving cell cycle factors (Figure 3). A more focused analysis revealed an enrichment, among our top barcoded RH-seq hits, for protein-protein interactions with one another as tabulated in BioGRID (Oughtred et al., 2021), to an extent beyond the null expectation (resampling $p = 0.014$). We also implemented qualitative gene set enrichment tests, which revealed that chromosome segregation and mitosis factors, although relatively few in number among our top barcoded RH-seq hit loci, were significantly enriched relative to the genomic null (Table 1). And we developed a complementary, quantitative test to screen Gene Ontology terms for large allelic effect size (the impact on thermotolerance when the *S. cerevisiae* allele of a given gene was disrupted in the hybrid, as a difference from the analogous quantity for the *S. paradoxus* allele; see Methods). Top-scoring in this test was a mitosis gene group, encoding components of the septin ring (GO:0000921; resampling $p < 0.0001$). Together, these results suggest that our top thermotolerance gene hits share commonalities in function, most notably involving cell cycle factors. This dovetails with previous phenotypic and genetic characterization of yeast thermotolerance, including the breakdown of cell division in heat-treated *S. paradoxus* (Weiss et al., 2018a), and supports a model in which *S. cerevisiae* acquired thermotolerance in part by resolving the latter cell cycle defect.

The genetics of yeast thermotolerance likely also involves mechanisms beside mitosis, given the known role of mitochondrial genes (Baker et al., 2019; Li et al., 2019) and those operating during stationary phase (AlZaben et al., 2021). Indeed, functional-genomic tests revealed enrichment for secretion genes and for regulatory factors in our top RH-seq hits, although no such group constituted a large proportion of the total hit list (Table 1). Annotations in transcription and translation, mitochondrial function, and signaling were also apparent in our top thermotolerance loci (Figure 2B). These trends are consistent with a scenario in which evolution built the trait in *S. cerevisiae* by tweaking an array of housekeeping mechanisms, beside those that involve cell cycle machinery.

Evolutionary analysis of thermotolerance genes

We anticipated that sequence analyses of the genes we had mapped to thermotolerance by barcoded RH-seq could shed light on the evolutionary history of the trait. To explore this, we used a phylogenetic approach in *Saccharomyces sensu stricto*. We first inferred species-specific branch lengths in the phylogeny of each gene in turn,

and focused on the lineage leading to *S. cerevisiae*. The distribution of branch lengths along this lineage among top thermotolerance gene hits was not detectably different from that of the genome as a whole, with the exception of two rapidly evolving thermotolerance genes, *TAF2* and *BUL1*, encoding a transcription initiation factor and ubiquitin ligase adapter respectively (Figure S5). Separately, we quantified protein evolutionary rates in top hits from barcoded RH-seq. A branch-site phylogenetic modeling approach (Yang, 2007) detected striking evidence for positive selection along the *S. cerevisiae* lineage in the amino acid permease *GNP1*, the kinetochore DNA binding factor *CBF2*, and the sister chromatid cohesion factor *CTF18* (Figure 4). Interestingly, however, McDonald-Kreitman tests (McDonald and Kreitman, 1991) on population-genomic data did not detect an overall excess of amino acid variation relative to synonymous changes, at these three genes or any other barcoded RH-seq hit locus (Table S7). Thus, even at genes harboring individual codons with likely signatures of selection, we could not detect evidence for a scenario where *S. cerevisiae* stacked up a large number of unique amino acid changes, in the evolution of thermotolerance. Together, however, our analyses do highlight thermotolerance genes with marked signal for derived alleles in *S. cerevisiae* at single codons or in the overall DNA sequence—cases where species divergence is likely to be of phenotypic and evolutionary importance.

DISCUSSION

RH-seq power and the interpretation of mapped loci

In this work, we established the barcoded RH-seq method for genetic dissection of trait variation between diverged lineages. RH-seq falls into a family of recently-developed methods that can dissect natural trait variation across species barriers (Weiss and Brem, 2019). A chief distinction of RH-seq is its low cost and low overhead, and the barcoding feature we add here cuts down labor and cost even further, enabling high replication.

Our application to yeast thermotolerance serves as an informative model for the performance of barcoded RH-seq on highly genetically complex traits. We pinpointed dozens of candidate genes at which species-level variation contributes to growth at high temperature. And yet we also observed evidence for a sizeable false negative rate among our barcoded RH-seq results, since some validated thermotolerance loci from our earlier screen did not appear among the hits here. Likewise, a separate barcoded RH-seq mapping of yeast species' differences in growth under milder heat stress revealed little signal above the noise (Table S6 and Table S8), likely reflecting very weak genetic effects under this condition. We thus expect that, as would be true for a classical linkage or association scan, the statistical power of a barcoded RH-seq experiment is a function of signal to-noise, genetic complexity, and genetic effect size; and that many thermotolerance loci remain to be identified even in our very deep set of screen results from high-temperature growth.

By virtue of our focus on pro-thermotolerance alleles in *S. cerevisiae*, our work has left open the functional and evolutionary genomics of loci at which the allele from *S. cerevisiae* instead conferred worse thermotolerance than that of *S. paradoxus*, when each in turn was uncovered in the hybrid. Our barcoded RH-seq identified a number of such genes at high statistical significance. These loci may well reflect the accumulation of advantageous alleles in *S. paradoxus*, or deleterious alleles in *S. cerevisiae*, by drift, even as the latter was under selection to improve the trait in evolutionary history. Analogously, in linkage mapping results, the effect of an allele in recombinant progeny from a cross often does not conform to that expected from the respective parent's phenotype (Brem and Kruglyak, 2005; Burke and Arnold, 2001). It is also possible that some such allelic effects are the product of epistatic interactions between a locus of interest and the hybrid background, and would be phenotypically buffered (and thus evolutionarily irrelevant) in the purebred species. Processes of compensatory evolution, or evolution under stabilizing selection could also explain the presence of these loci where the barcoded RH-seq data suggest a thermotolerance advantage might stem from the *S. paradoxus* allele. Molecular validation will be necessary to confirm the phenotypic impact of variation at our mapped loci, and its potential dependence on genetic background.

That said, we consider genes with pro-thermotolerance *S. cerevisiae* alleles according to barcoded RH-seq to be strong candidates for *bona fide* determinants of the trait from the wild in this species. Indeed, earlier work has shown that for such genes mapped by RH-seq in the hybrid, the advantage of *S. cerevisiae* alleles is borne out in tests in purebred backgrounds (Weiss et al., 2018a). Accordingly, we have shown here that as a cohort, barcoded RH-seq hits with advantageous *S. cerevisiae* alleles exhibit functional and sequence-based attributes consistent with a role in thermotolerance evolution in the wild.

Cellular and molecular mechanisms of thermotolerance

Our top RH-seq hits revealed strong evidence for chromosome segregation and other mitosis functions as a linchpin of *S. cerevisiae* thermotolerance. As a complement to earlier characterization of six such genes (*APC1, ESP1, DYN1, MYO1, CEP3,* and *SCC2*) (Abrams et al., 2021a; Weiss et al., 2018a), we now report seven new thermotolerance determinants that function in cell division (*MEC1, MLH1, CTF13, CTF18, MCM21, CBF2,* and *MYO2*). The emerging picture is one in which the ancestor of modern-day *S. cerevisiae*, faced with dysfunction of a slew of mitotic factors at high temperature, acquired variants across the genome to shore up their activity under these conditions. Under one model of *S. cerevisiae* evolution, the particular niche to which this species specialized was one of avid fermentation, producing (and resisting) heat and ethanol at levels that eliminated its microbial competitors (Goddard, 2008; Salvadó et al., 2011). In such a scenario, the maximum benefit could well accrue to the organism if it were able to undergo rapid cell division under the challenging conditions of its own making. Consistent with this notion, another budding yeast, *Hanseniaspora,* which often dominates in early fermentation prior to takeover by *S. cerevisiae* (Fleet, 2003), underwent evolutionary loss of much of the cell-cycle checkpoint machinery, consistent

with a strategy of accelerated growth at any cost to outcompete other species at the respective stage (Steenwyk et al., 2019).

However, since our current hit list includes many genes from other housekeeping pathways, from transcription/translation to transport and lipid metabolism, mitosis does not appear to be the whole mechanistic story for the thermotolerance trait in *S. cerevisiae*. Indeed, other housekeeping factors also showed up in our previous screen (Weiss et al., 2018a) and in an elegant complementary study of mitochondrial determinants of thermotolerance divergence between yeast species (Baker et al., 2019; Li et al., 2019). The panoply of functions detected among our mapped loci conforms well to current models of the mechanisms of thermotolerance, which invoke many essential genes and housekeeping processes (Leuenberger et al., 2017).

The latter idea emerged largely from a proteomic study which showed that thermotolerant organisms had higher thermostability of essential proteins of many functions, across the tree of life (Leuenberger et al., 2017). Were sequence changes that led to improved protein stability a linchpin of thermotolerance evolution in *S. cerevisiae*? Our data are consistent with a mechanistic role for properties of the protein sequences of many thermotolerance genes, in that variation in coding regions has come to the fore in our sequence tests here and those of an earlier small-scale analysis (Abrams et al., 2021a). And interestingly, an experimental case study of one of our mapped thermotolerance loci revealed no impact on the trait from variation in the promoter, only from that in the coding region (Abrams et al., 2021a). We cannot rule out noncoding determinants in some cases, especially given that a few hundred genes exhibit temperature-dependent *cis*-regulatory programs unique to *S. cerevisiae* (Li and Fay, 2017; Tirosh et al., 2009). But if coding regions do hold the exclusive key to the mechanism of *S. cerevisiae* thermotolerance, they could well involve variants that improve protein function and regulation alongside folding/structure at high temperature. Overall, then, we envision that nature could have used a variety of molecular mechanisms in building the trait, given the apparent complexity of the problem. Biochemical studies will be necessary to nail down exactly how *S. cerevisiae* alleles advance thermotolerance.

In summary, our data reveal a newly detailed picture of the highly polygenic architecture for a natural trait divergence between species. It is tempting to speculate that evolution may draw on a vast number of variants across the genome to refine a trait over millions of generations, making effects stronger, weaker, or less pleiotropic, adding regulatory control, and so on (Orr, 1998). If so, these architectures may ultimately conform to the omnigenic model (Boyle et al., 2017)–which was originally applied to human disease genetics, but may also prove to be an apt description of ancient adaptations.

DATA AVAILABILITY

Sequencing data are deposited in the Sequence Read Archive under the accession PRJNA735401. Strains and plasmids are available upon request. Custom scripts for the

barcoded RH-seq analysis are available at https://github.com/melanieabrams-pub/RH-seq_with_barcoding.

ACKNOWLEDGEMENTS

The authors thank Adam Arkin for his generosity with computational resources; Morgan Price, Lori Huberman, and members of the John Dueber lab for discussions about barcoded transposon mutagenesis; and sequencing; and Abel Duarte for cloning advice. This work was supported by NSF GRFP DGE 1752814 to M.A. and NIH R01 GM120430 to R.B.B.

REFERENCES

Abrams, M.B., Dubin, C.A., AlZaben, F., Bravo, J., Joubert, P.M., Weiss, C.V., and Brem, R.B. (2021). Population and comparative genetics of thermotolerance divergence between yeast species. G3 Genes|Genomes|Genetics.

AlZaben, F., Chuong, J.N., Abrams, M.B., and Brem, R.B. (2021). Joint effects of genes underlying a temperature specialization tradeoff in yeast. PLoS Genet 17, e1009793.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29.

Baker, E.P., Peris, D., Moriarty, R.V., Li, X.C., Fay, J.C., and Hittinger, C.T. (2019). Mitochondrial DNA and temperature tolerance in lager yeasts. Sci Adv 5, eaav1869.

Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177–1186.

Brem, R.B., and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A 102, 1572–1577.

Burke, J.M., and Arnold, M.L. (2001). Genetics and the fitness of hybrids. Annu Rev Genet 35, 31–52.

Coradetti, S.T., Pinel, D., Geiselman, G.M., Ito, M., Mondo, S.J., Reilly, M.C., Cheng, Y.-F., Bauer, S., Grigoriev, I.V., Gladden, J.M., et al. (2018). Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. ELife 7, e32110.

Dubin, C.A., Roop, J.I., and Brem, R.B. (2020). Divergence of Peroxisome Membrane Gene Sequence and Expression Between Yeast Species. G3: Genes, Genomes, Genetics 10, 2079–2085.

Fleet, G.H. (2003). Yeast interactions and wine flavour. International Journal of Food Microbiology 86, 11–22.

Goddard, M.R. (2008). Quantifying the complexities of Saccharomyces cerevisiae's ecosystem engineering via fermentation. Ecology 89, 2077–2082.

Gonçalves, P., Valério, E., Correia, C., Almeida, J.M.G.C.F. de, and Sampaio, J.P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric Saccharomyces Species. PLOS ONE 6, e20739.

Hittinger, C.T. (2013). Saccharomyces diversity and evolution: a budding model genus. Trends Genet 29, 309–317.
Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth. Biol. 4, 975–986.
Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P.J., von Mering, C., Claassen, M., and Picotti, P. (2017). Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science 355.
Li, X.C., and Fay, J.C. (2017). Cis-Regulatory Divergence in Gene Expression between Two Thermally Divergent Yeast Species. Genome Biol Evol 9, 1120–1129.
Li, X.C., Peris, D., Hittinger, C.T., Sia, E.A., and Fay, J.C. (2019). Mitochondria-encoded genes contribute to evolution of heat and cold tolerance in yeast. Sci. Adv. 5, eaav1848.
McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654.
Orr, H.A. (1998). The Population Genetics of Adaptation: The Distribution of Factors Fixed During Adaptive Evolution. Evolution 52, 935–949.
Ott, J., Wang, J., and Leal, S.M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet 16, 275–284.
Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.-J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci 30, 187–200.
Peter, J., Chiara, M.D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344.
Salvadó, Z., Arroyo-López, F.N., Guillamón, J.M., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus Saccharomyces. Appl Environ Microbiol 77, 2292–2302.
Steenwyk, J.L., Opulente, D.A., Kominek, J., Shen, X.-X., Zhou, X., Labella, A.L., Bradley, N.P., Eichman, B.F., Čadež, N., Libkind, D., et al. (2019). Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. PLOS Biology 17, e3000255.
Stern, D.L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. Trends Genet 30, 547–554.
Sweeney, J.Y., Kuehne, H.A., and Sniegowski, P.D. (2004). Sympatric natural Saccharomyces cerevisiae and S. paradoxus populations have different thermal growth profiles. FEMS Yeast Res. 4, 521–525.
Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research 49, D605–D612.
Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat Rev Genet 20, 467–484.

Tirosh, I., Reikhav, S., Levy, A.A., and Barkai, N. (2009). A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. Science 324, 659–662.
Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191.
Weiss, C.V., and Brem, R.B. (2019). Dissecting Trait Variation across Species Barriers. Trends in Ecology & Evolution 34, 1131–1136.
Weiss, C.V., Roop, J.I., Hackley, R.K., Chuong, J.N., Grigoriev, I.V., Arkin, A.P., Skerker, J.M., and Brem, R.B. (2018). Genetic dissection of interspecific differences in yeast thermotolerance. Nature Genetics 50, 1501.
Wetmore, K.M., Price, M.N., Waters, R.J., Lamson, J.S., He, J., Hoover, C.A., Blow, M.J., Bristow, J., Butland, G., Arkin, A.P., et al. (2015). Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. MBio 6, e00306-15.
Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science 285, 901–906.
Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24, 1586–1591.

FIGURES

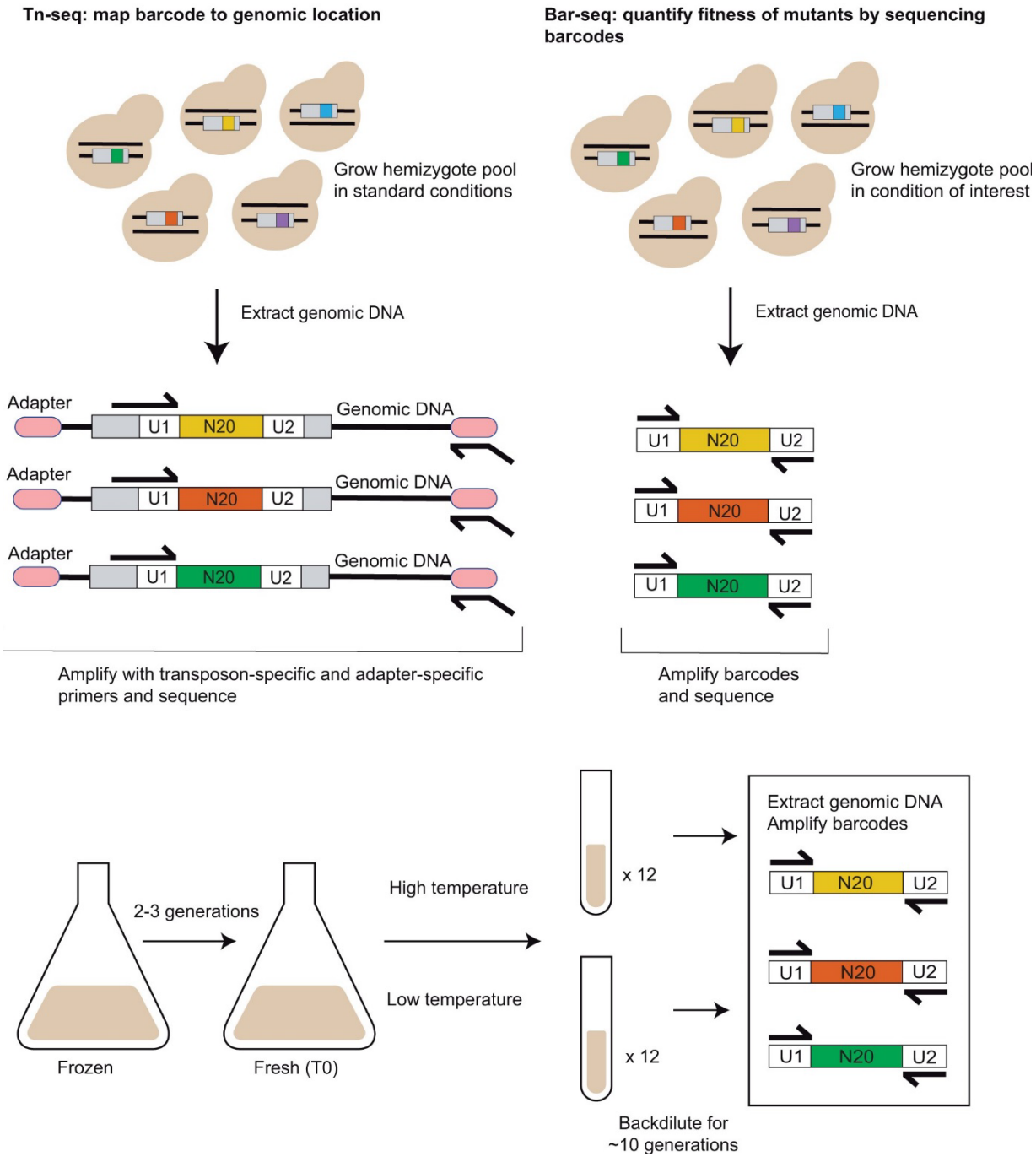
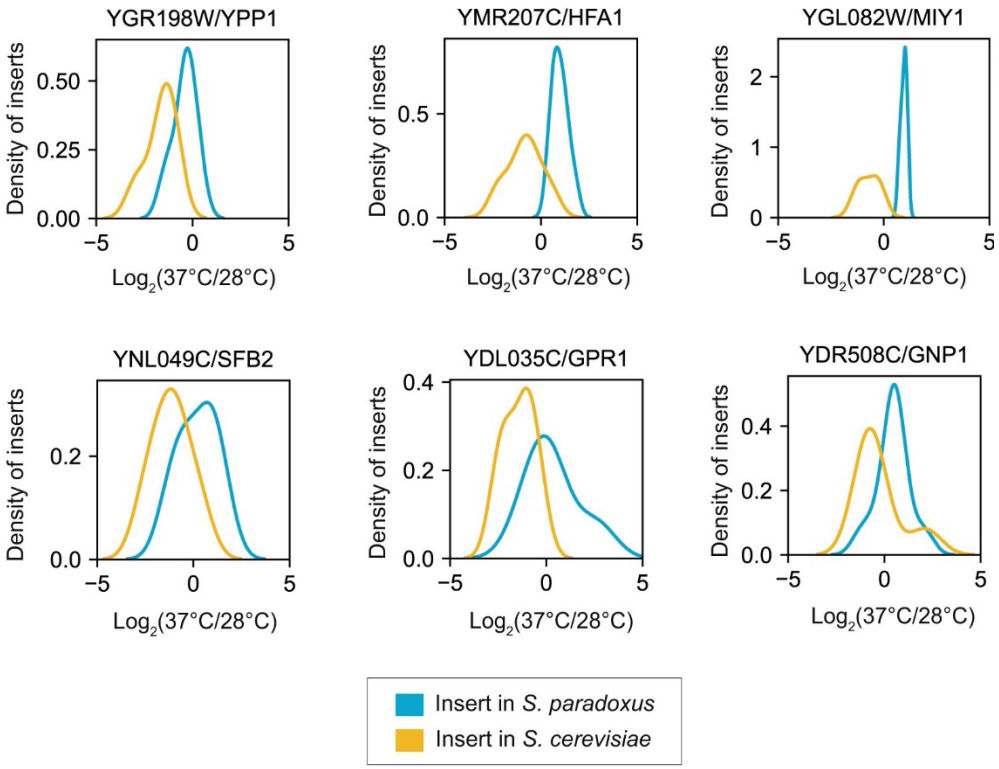


Figure 1. Barcoded RH-seq mapping of yeast thermotolerance loci. (A) Barcoded RH-seq sequencing analysis steps. Left, in a pool of *S. cerevisiae* x *S. paradoxus* hybrid hemizygotes, each harboring a transposon (grey rectangle) marked with a unique 20-mer barcode (multicolored) flanked by universal primer sites (U1 and U2), each barcode

is associated with its insertion location by transposon sequencing (Tn-seq). Genomic DNA from the pool is extracted, sheared, and ligated to universal adapters (pink ovals), followed by PCR amplification with a transposon-specific primer (forward black arrow) and an adapter-specific primer (reverse black arrow) and sequencing. Right, for barcode sequencing (Bar-seq) to quantify hemizygote strain abundance after pool growth in a condition of interest, genomic DNA is used as input to PCR with primers to universal primer sites for sequencing. (B) Thermotolerance RH-seq screen design. An aliquot of the hemizygote pool was thawed and cultured in large format, then split into small replicate cultures, each maintained in logarithmic growth phase at the temperature of interest by back-dilution, followed by quantification by Bar-seq.

A.


YGR198W/YPP1
YMR207C/HFA1
YGL082W/MIY1
YNL049C/SFB2
YDL035C/GPR1
YDR508C/GNP1
Density of inserts
Log2(37°C/28°C)
Insert in S. paradoxus
Insert in S. cerevisiae

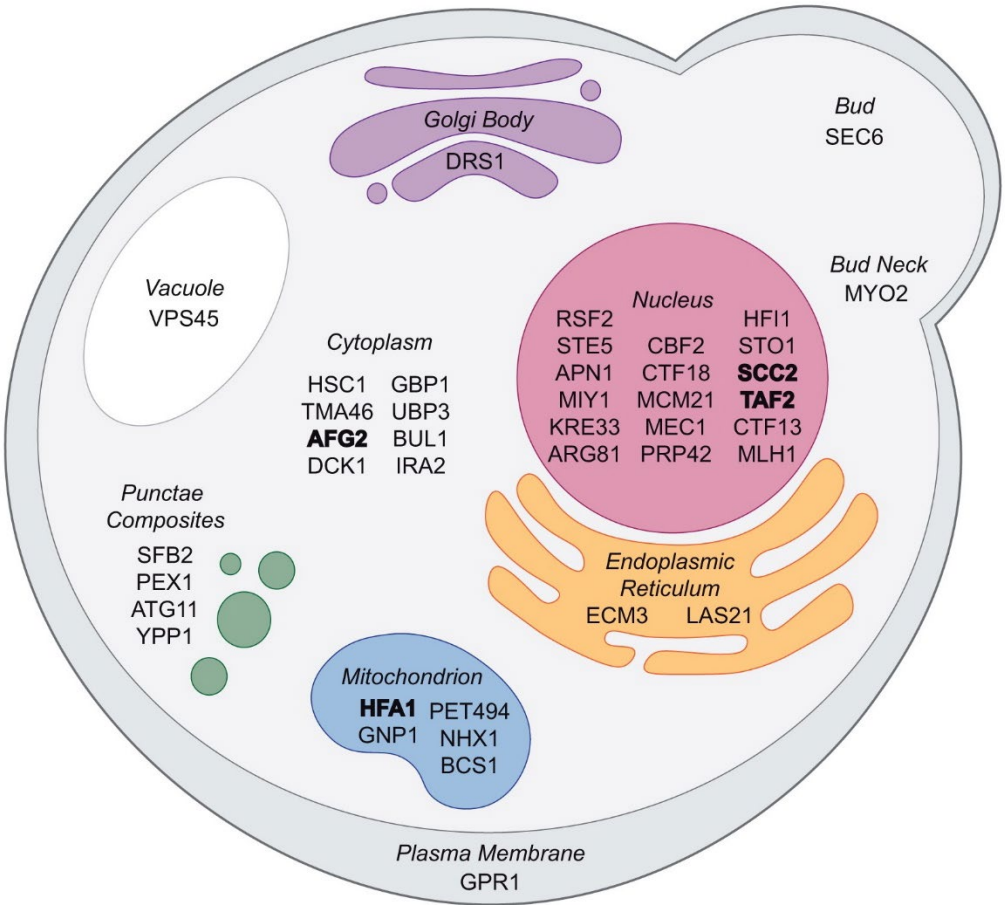
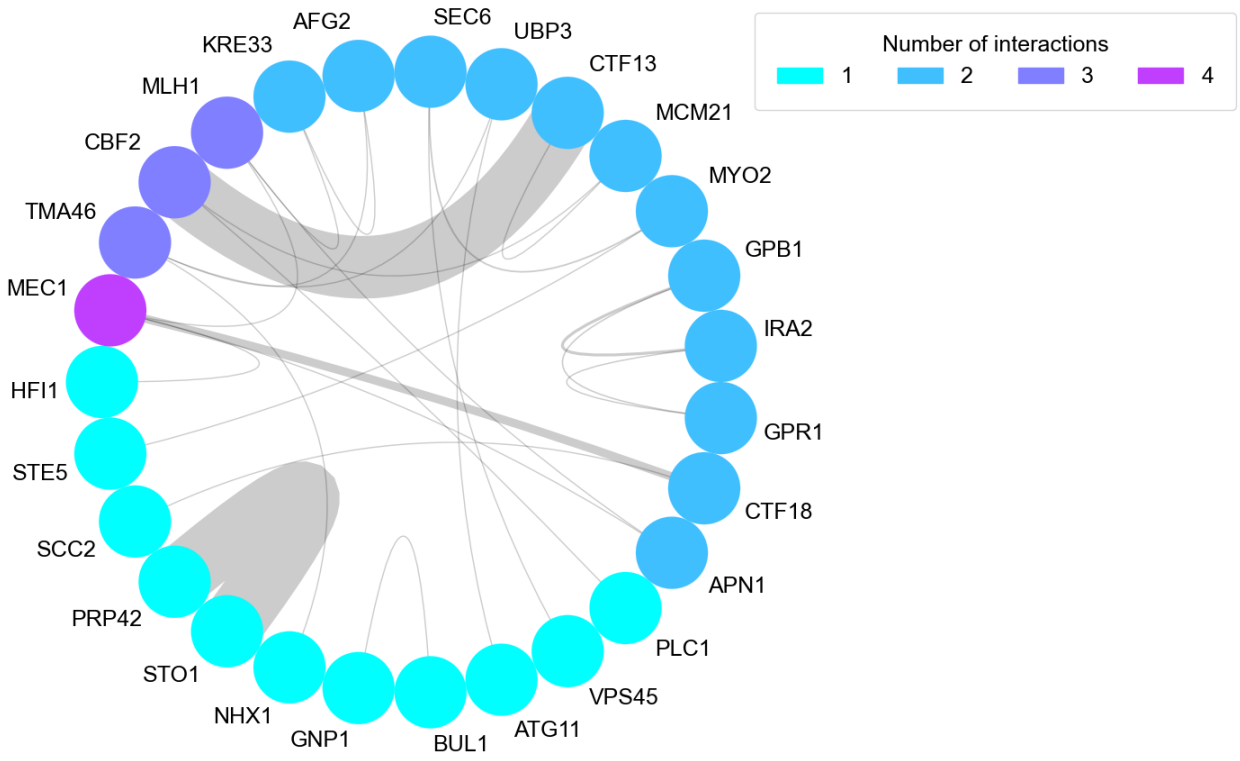
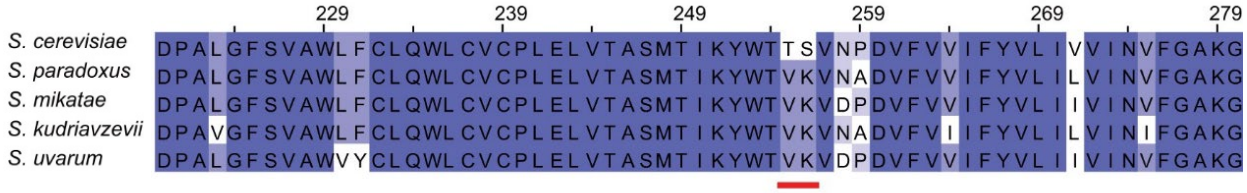


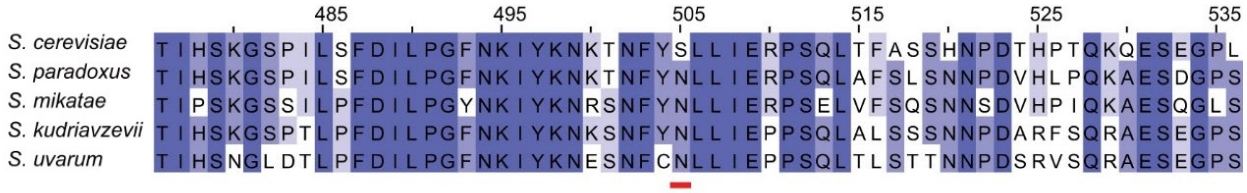
Figure 2. Hits from barcoded RH-seq mapping of yeast thermotolerance. (A) Each panel reports barcoded RH-seq results for a gene at which the *S. cerevisiae* allele is associated with better thermotolerance than the *S. paradoxus* allele, when uncovered in the hybrid background. In a given panel, the *x*-axis reports the $\log_2$ of abundance, measured by RH-seq after selection at 37°C, of a clone harboring a barcoded transposon insertion in the indicated species' allele in a given replicate, as a difference from the analogous quantity for that clone after selection at 28°C on average across replicates. The *y*-axis reports the proportion of observations of all clones bearing insertions in the indicated allele that exhibited the abundance ratio on the *x*, as a kernel density estimate. Shown are the top six genes from among all barcoded RH-seq hit loci in terms of allelic effect size. (B) Subcellular localization of RH-seq hit genes, where available from (Pierleoni et al., 2007) and (Huh et al., 2003). Genes at which effects of allelic variation on thermotolerance were reported previously (Li et al., 2019; Weiss et al., 2018a) are denoted in bold type.


Number of interactions
1
2
3
4
AFG2
SEC6
UBP3
KRE33
CTF13
MLH1
MCM21
CBF2
MYO2
TMA46
GPB1
MEC1
IRA2
HFI1
GPR1
STE5
CTF18
SCC2
APN1
PRP42
PLC1
STO1
VPS45
NHX1
ATG11
GNP1
BUL1

A.



B.



C.

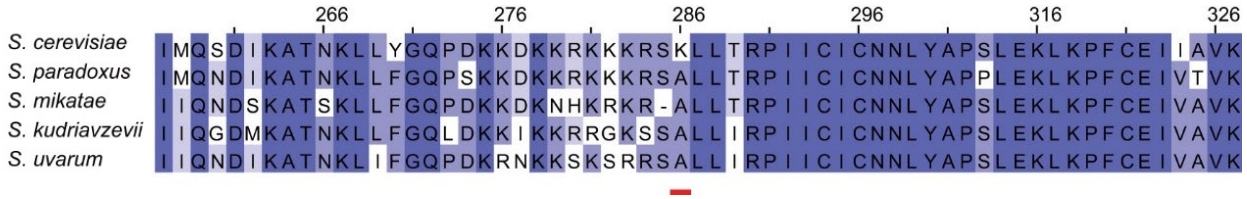


Figure 4. Codons under positive selection in thermotolerance loci. Each panel shows the amino acid sequence context of the codon(s) (red bar) inferred to be under positive selection along the *S. cerevisiae* lineage, in a hit gene from RH-seq thermotolerance mapping. Alignments are colored by percent identity, with darker purples indicating a higher percent identity. (A) *YDR508C/GNP1*. (B) *YGR140W/CBF2*. (C) *YMR078C/CTF18*.

TABLES

GO term	$n_{observed}$: $n_{expected}$	Adjusted *p*	Name	Total n
Cellular Component				
GO:0000775	5:1	0.0366	chromosome, centromeric region	75
GO:0000778	4:0	0.0256	kinetochore	40
Molecular Function				
GO:0000149	3:0	0.0701	SNARE binding	28
GO:0008081	2:0	0.0256	phosphoric diester hydrolase activity	11
GO:0004843	2:0	0.0998	thiol-dependent deubiquitinase	24
Biological Process				
GO:0007165	3:0	0.0923	signal transduction	59
GO:0001403	3:0	0.0923	invasive growth in response to glucose limitation	42
GO:0046580	2:0	0.0256	negative regulation of Ras protein signal transduction	6
GO:0001934	2:0	0.0256	positive regulation of protein phosphorylation	5
GO:0016042	2:0	0.0923	lipid catabolic process	26
GO:0034087	2:0	0.0923	establishment of mitotic sister chromatid cohesion	16

Table 1. Functional enrichment among thermotolerance loci. Each row with numerical data reports a Gene Ontology (GO) term enriched for RH-seq hit genes. $n_{observed}$, the number of genes from among top hits from thermotolerance RH-seq that were annotated with the term. $n_{expected}$, the number of genes annotated with the term in the same number of randomly chosen genes from the genome, as a median across samples. Adjusted *p*, resampling-based significance of the enrichment after Benjamini-Hochberg

Oligo name	Sequence (5' to 3')	Index Identifier (if applicable)	Index Sequence	Notes
Barcoding cloning				
Random Barcodes U1 — N20 — U2	GATGTCCACG AGGTCTCTNN NNNNNNNNNN NNNNNNNNCG TACGCTGCAG GTCGAC			Random barcodes flanked by universal primer sites
FW_BbsI_JC	TCACACAAGT TTGTACAAAAA AGCAGGCTGG AGCTCGgaaga cATCCCTGAT GTCCACGAGG TCTCT			Forward primer to amplify barcodes for cloning into vector
REV_BbsI_JC	CTCAACCACT TTGTACAAGA AAGCTGGGTG GATCCgaagac CGCGTTGTCG ACCTGCAGCG TACG			Reverse primer for to amplify barcodes for cloning into vector
Tn-seq				
JC8 Tn-specific FW Primer	ATGATACGGC GACCACCGAG ATCTACACTCT TTCCCTACAC GACGCTCTTC CGATCTNNNN NNCCCTGATG TCCACGAGGT CTCT			Forward primer to sequence transposon insertions; homologous to transposon (Figure 1A, left)
P7_MOD_TS_index1	CAAGCAGAAG ACGGCATACG AGATCGTGAT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCT	Index1	ATCACG	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index2	CAAGCAGAAG ACGGCATACG	Index2	CGATGT	Indexed reverse primer to

	AGATACATCG GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCT			sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index3	CAAGCAGAAG ACGGCATACG AGATGCCTAA GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCT	Index3	TTAGGC	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index4	CAAGCAGAAG ACGGCATACG AGATTGGTCA GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCT	Index4	TGACCA	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index5	CAAGCAGAAG ACGGCATACG AGATCACTGT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCT	Index5	ACAGTG	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index6	CAAGCAGAAG ACGGCATACG AGATATTGGC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCT	Index6	GCCAAT	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index7	CAAGCAGAAG ACGGCATACG AGATGATCTG GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCT	Index7	CAGATC	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)

P7_MOD_TS_index8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index8	ACTTGA	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
Bar-seq				
P1_BS3_IT001	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNGCACTAGTCGACCTGCAGCGTACG	IT001	ATCACG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT002	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNTGTAGCGTCGACCTGCAGCGTACG	IT002	CGATGT	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT003	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCGGATTGTCGACCTGCAGCGTACG	IT003	TTAGGC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT004	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCAGTGTCGACCTGCAGCGTACG	IT004	TGACCA	Indexed forward Bar-seq primer (Figure 1A, right)

P1_BS3_IT005	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNG TGACAGTCGA CCTGCAGCGT ACG	IT005	ACAGTG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT006	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN TAACCGGTCG ACCTGCAGCG TACG	IT006	GCCAAT	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT007	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NCTAGACGTC GACCTGCAGC GTACG	IT007	CAGATC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT008	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NNAGTTCAGT CGACCTGCAG CGTACG	IT008	ACTTGA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT009	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNG ACTAGGTCGA CCTGCAGCGT ACG	IT009	GATCAG	Indexed forward Bar-seq primer (Figure 1A, right)

P1_BS3_IT010	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN TTCGATGTCG ACCTGCAGCG TACG	IT010	TAGCTT	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT011	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NCATCGGGTC GACCTGCAGC GTACG	IT011	GGCTAC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT012	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NNATGTTCGT CGACCTGCAG CGTACG	IT012	CTTGTA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT013	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNA ACTGAGTCGA CCTGCAGCGT ACG	IT013	AGTCAA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT014	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN CCTTGAGTCG ACCTGCAGCG TACG	IT014	AGTTCC	Indexed forward Bar-seq primer (Figure 1A, right)

P1_BS3_IT015	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NACTGTAGTC GACCTGCAGC GTACG	IT015	ATGTCA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT016	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NNCCTGCCGT CGACCTGCAG CGTACG	IT016	CCGTCC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT017	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNG AGATGGTCGA CCTGCAGCGT ACG	IT017	GTAGAG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT018	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN CGCCTGGTCG ACCTGCAGCG TACG	IT018	GTCCGC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT019	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NAAAGTGGTC GACCTGCAGC GTACG	IT019	GTGAAA	Indexed forward Bar-seq primer (Figure 1A, right)

P1_BS3_IT020	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NNCCGGTGGT CGACCTGCAG CGTACG	IT020	GTGGCC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT021	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNG CTTTGGTCGA CCTGCAGCGT ACG	IT021	GTTTCG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT022	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN GCATGCGTCG ACCTGCAGCG TACG	IT022	CGTACG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT023	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NGGTGAGGTC GACCTGCAGC GTACG	IT023	GAGTGG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT024	AATGATACGG CGACCACCGA GATCTACACT CTTTCCCTAC ACGACGCTCT TCCGATCTNN NNCGATGGGT CGACCTGCAG CGTACG	IT024	GGTAGC	Indexed forward Bar-seq primer (Figure 1A, right)

P2_BS3_IT001	CAAGCAGAAG ACGGCATACG AGATCGTGAT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT001	ATCACG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT002	CAAGCAGAAG ACGGCATACG AGATACATCG GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT002	CGATGT	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT003	CAAGCAGAAG ACGGCATACG AGATGCCTAA GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT003	TTAGGC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT004	CAAGCAGAAG ACGGCATACG AGATTGGTCA GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT004	TGACCA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT005	CAAGCAGAAG ACGGCATACG AGATCACTGT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT005	ACAGTG	Indexed reverse Bar-seq primer (Figure 1A, right)

P2_BS3_IT006	CAAGCAGAAG ACGGCATACG AGATATTGGC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT006	GCCAAT	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT007	CAAGCAGAAG ACGGCATACG AGATGATCTG GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT007	CAGATC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT008	CAAGCAGAAG ACGGCATACG AGATTCAAGT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT008	ACTTGA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT009	CAAGCAGAAG ACGGCATACG AGATCTGATC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT009	GATCAG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT010	CAAGCAGAAG ACGGCATACG AGATAAGCTA GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT010	TAGCTT	Indexed reverse Bar-seq primer (Figure 1A, right)

P2_BS3_IT011	CAAGCAGAAG ACGGCATACG AGATGTAGCC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT011	GGCTAC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT012	CAAGCAGAAG ACGGCATACG AGATTACAAG GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT012	CTTGTA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT013	CAAGCAGAAG ACGGCATACG AGATTTGACT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT013	AGTCAA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT014	CAAGCAGAAG ACGGCATACG AGATGGAACT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT014	AGTTCC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT015	CAAGCAGAAG ACGGCATACG AGATTGACAT GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT015	ATGTCA	Indexed reverse Bar-seq primer (Figure 1A, right)

P2_BS3_IT016	CAAGCAGAAG ACGGCATACG AGATGGACGG GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT016	CCGTCC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT017	CAAGCAGAAG ACGGCATACG AGATCTCTAC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT017	GTAGAG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT018	CAAGCAGAAG ACGGCATACG AGATGCGGAC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT018	GTCCGC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT019	CAAGCAGAAG ACGGCATACG AGATTTTCAC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT019	GTGAAA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT020	CAAGCAGAAG ACGGCATACG AGATGGCCAC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT020	GTGGCC	Indexed reverse Bar-seq primer (Figure 1A, right)

P2_BS3_IT021	CAAGCAGAAG ACGGCATACG AGATCGAAAC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT021	GTTTCG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT022	CAAGCAGAAG ACGGCATACG AGATCGTACG GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT022	CGTACG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT023	CAAGCAGAAG ACGGCATACG AGATCCACTC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT023	GAGTGG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT024	CAAGCAGAAG ACGGCATACG AGATGCTACC GTGACTGGAG TTCAGACGTG TGCTCTTCCG ATCTGATGTC CACGAGGTCT CT	IT024	GGTAGC	Indexed reverse Bar-seq primer (Figure 1A, right)

Supplementary Table 2. Oligonucleotides used in this study.

Experiment	Library	Reads
E. coli vector pool Bar-seq	RBJC009_IT013	93,277,156
	RBJC009_IT014	73,914,499
	RBJC009_IT015	79,273,122
	RBJC009_IT016	81,404,878
	RBJC45_IT013	100,794,463
	RBJC45_IT014	92,410,111
	RBJC45_IT015	91,931,624
	RBJC45_I0T16	96,631,037
	RBJC009_IT013	93,277,156
	RBJC009_IT014	73,914,499
	RBJC009_IT015	79,273,122
	RBJC45_IT016	96,631,037
37°C vs. 28°C barcoded RH-seq in yeast	RMBA038A_IT001	16,042,537
	RBMA038A_IT002	19,451,725
	RBMA038A_IT003	16,313,279
	RBMA038A_IT004	19,512,043
	RBMA038A_IT005	13,625,004
	RBMA038A_IT006	22,257,379
	RBMA038A_IT007	22,385,133
	RBMA038A_IT008	16,359,601
	RBMA038A_IT009	20,118,293
	RBMA038A_IT010	19,121,369
	RBMA038A_IT011	18,887,044
	RBMA038A_IT012	15,895,905
	RBMA038A_IT013	18,922,229
	RBMA038A_IT014	26,602
	RBMA038A_IT015	16,571,744
	RBMA038A_IT016	18,783,445
	RBMA038A_IT017	17,671,257
	RBMA038A_IT018	18,040,605
	RBMA038A_IT019	16,929,985
	RBMA038A_IT020	15,319,721
	RBMA038A_IT021	16,897,401
	RBMA038A_IT022	16,372,841
	RBMA038A_IT023	16,679,583
	RBMA038A_IT024	13,875,240
36°C vs. 28°C barcoded RH-seq in yeast	RMBA039A_IT001	28,004,341
	RBMA039A_IT002	19,283,097
	RBMA039A_IT003	12,427,040
	RBMA039A_IT004	13,934,236
	RBMA039A_IT005	12,068,707
	RBMA039A_IT006	60,072,532

	RBMA039A_IT008	12,500,340
	RBMA039A_IT009	25,029,516
	RBMA039A_IT010	14,765,007
	RBMA039A_IT011	7,059,133
	RBMA039A_IT012	19,837,690
	RBMA039A_IT013	14,393,602
	RBMA039A_IT014	19,572,649
	RBMA039A_IT015	8,814,383
	RBMA039A_IT016	14,587,181
	RBMA039A_IT017	8,619,661
	RBMA039A_IT018	13,216,843
	RBMA039A_IT019	12,499,486
	RBMA039A_IT020	11,963,866
	RBMA039A_IT021	6,626,989
	RBMA039A_IT022	11,649,542
	RBMA039A_IT023	5,813,357
	RBMA039A_IT024	9,349,172

Supplementary Table 3. Bar-seq sequencing data sets. Each row reports numbers of reads sequenced for the indicated Bar-seq experiment. The first set of rows reports results from a check of barcoded piggyBac transposon plasmids as in Figure S1C; the remaining rows report results from quantification of yeast hemizygote insertion genotypes after competition in the indicated condition, as in Figure 1B of the main text. Experiment identifiers are from BioProject PRJNA735401.

Pool	Library	Reads	Platform	Facility
67	RBJC37	38,713,102	Novaseq SP PE150	UC Berkeley
69	RBJC38	38,875,221		
69	RBJC39	43,194,450		
69	RBJC40	39,778,862		
69	RBJC41	38,836,065		
67	RBJC42	39,265,466		
67	RBJC43	47,124,575		
67	RBJC44	39,762,187		
67	RBJC48	91,531,071	HiSeq4000 PE150	Novogene
67	RBJC48_reseq	86,892,060		
70	RBCJ51	86,254,426		
70	RBCJ51_reseq	86,130,880		
70	RBJC52	52,108,306		
70	RBJC52_reseq	53,363,169		
71	RBJC54	88,154,532		
71	RBJC54_reseq	86,878,835		
71	RBJC55	90,265,981		
71	RBJC55_reseq	82,130,170		
69	RBJC57	84,296,399		
69	RBJC57_reseq	85,606,080		

Supplementary Table 4. Tn-seq sequencing data sets. Each row reports numbers of reads from the indicated sequencing of insertion positions of barcoded transposons in the *S. cerevisiae* x *S. paradoxus* hybrid, as in Figure 1A, left, of the main text. Experiment identifiers are from PRJNA735401; “reseq” indicates the reads from a technical replicate performed to gather additional reads for the indicated library.

Hit	Description
YGR198W/YPP1	Cargo-transport protein involved in endocytosis; interacts with phosphatidylinositol-4-kinase Stt4; GFP-fusion protein localizes to the cytoplasm; YGR198W is an essential gene
YMR207C/HFA1	Mitochondrial acetyl-coenzyme A carboxylase, catalyzes the production of malonyl-CoA in mitochondrial fatty acid biosynthesis
YGL082W/MIY1	Putative protein of unknown function; predicted prenylation/proteolysis target of Afc1p and Rce1p; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; YGL082W is not an essential gene
YNL049C/SFB2	Component of the Sec23p-Sfb2p heterodimer of the COPII vesicle coat, required for cargo selection during vesicle formation in ER to Golgi transport; homologous to Sec24p and Sfb3p
YDL035C/GPR1	Plasma membrane G protein coupled receptor (GPCR) that interacts with the heterotrimeric G protein alpha subunit, Gpa2p, and with Plc1p; sensor that integrates nutritional signals with the modulation of cell fate via PKA and cAMP synthesis
YDR508C/GNP1	High-affinity glutamine permease, also transports Leu, Ser, Thr, Cys, Met and Asn; expression is fully dependent on Grr1p and modulated by the Ssy1p-Ptr3p-Ssy5p (SPS) sensor of extracellular amino acids
YBR136W/MEC1	Genome integrity checkpoint protein and PI kinase superfamily member; signal transducer required for cell cycle arrest and transcriptional responses prompted by damaged or unreplicated DNA; monitors and participates in meiotic recombination
YML099C/ARG81	Zinc-finger transcription factor of the Zn(2)-Cys(6) binuclear cluster domain type, involved in the regulation of arginine-responsive genes; acts with Arg80p and Arg82p
YPL254W/HFI1	Adaptor protein required for structural integrity of the SAGA complex, a histone acetyltransferase-coactivator complex that is involved in global regulation of gene expression through acetylation and transcription functions
YIL152W/VPR1	Putative protein of unknown function
YKL017C/HCS1	Hexameric DNA polymerase alpha-associated DNA helicase A involved in lagging strand DNA synthesis; contains single-stranded DNA stimulated ATPase and dATPase activities; replication protein A stimulates helicase and ATPase activities
YGR140W/CBF2	Essential kinetochore protein, component of the CBF3 multisubunit complex that binds to the CDEIII region of the centromere; Cbf2p also binds to the CDEII region possibly forming a different multimeric complex, ubiquitinated in vivo

YJR127C/RSF2	Zinc-finger protein involved in transcriptional control of both nuclear and mitochondrial genes, many of which specify products required for glycerol-based growth, respiration, and other functions
YDR375C/BCS1	Mitochondrial protein of the AAA ATPase family; has ATP-dependent chaperone activity; required for assembly of Rip1p and Qcr10p into cytochrome bc(1) complex; mutations in human homolog BCS1L are linked to neonatal mitochondrial diseases
YOR091W/TMA46	Protein of unknown function that associates with translating ribosomes; interacts with GTPase Rbg1p
YLR397C/AFG2	ATPase of the CDC48/PAS1/SEC18 (AAA) family, forms a hexameric complex; is essential for pre-60S maturation and release of several preribosome maturation factors; may be involved in degradation of aberrant mRNAs
YNL132W/KRE33	Essential protein, required for biogenesis of the small ribosomal subunit; heterozygous mutant shows haploinsufficiency in K1 killer toxin resistance
YMR078C/CTF18	Subunit of a complex with Ctf8p that shares some subunits with Replication Factor C and is required for sister chromatid cohesion; may have overlapping functions with Rad24p in the DNA damage replication checkpoint
YLR422W/DCK1	Protein of unknown function with similarity to human DOCK proteins (guanine nucleotide exchange factors); interacts with Ino4p; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm, YLR422W is not an essential protein
YMR125W/STO1	Large subunit of the nuclear mRNA cap-binding protein complex, interacts with Npl3p to carry nuclear poly(A)+ mRNA to cytoplasm; also involved in nuclear mRNA degradation and telomere maintenance; orthologous to mammalian CBP80
YOR371C/GPB1	Multistep regulator of cAMP-PKA signaling; inhibits PKA downstream of Gpa2p and Cyr1p, thereby increasing cAMP dependency; promotes ubiquitin-dependent proteolysis of Ira2p; regulated by G-alpha protein Gpa2p; homolog of Gpb2p
YMR094W/CTF13	Subunit of the CBF3 complex, which binds to the CDE III element of centromeres, bending the DNA upon binding, and may be involved in sister chromatid cohesion during mitosis
YMR167W/MLH1	Protein required for mismatch repair in mitosis and meiosis as well as crossing over during meiosis; forms a complex with Pms1p and Msh2p-Msh3p during mismatch repair; human homolog is associated with hereditary non-polyposis colon cancer
YDR103W/STE5	Pheromone-response scaffold protein that controls the mating decision; binds Ste11p, Ste7p, and Fus3p kinases, forming a MAPK cascade complex that interacts with the plasma membrane and Ste4p-Ste18p; allosteric activator of Fus3p

YDR318W/MCM21	Protein involved in minichromosome maintenance; component of the COMA complex (Ctf19p, Okp1p, Mcm21p, Ame1p) that bridges kinetochore subunits that are in contact with centromeric DNA and the subunits bound to microtubules
YAL026C/DRS2	Aminophospholipid translocase (flippase) that maintains membrane lipid asymmetry in post-Golgi secretory vesicles; contributes to clathrin-coated vesicle formation and endocytosis; mutations in human homolog ATP8B1 result in liver disease
YDR180W/SCC2	Subunit of cohesin loading factor (Scc2p-Scc4p), a complex required for loading of cohesin complexes onto chromosomes; involved in establishing sister chromatid cohesion during DSB repair via histone H2AX; evolutionarily-conserved adherin
YOR092W/ECM3	Non-essential protein of unknown function; involved in signal transduction and the genotoxic response; induced rapidly in response to treatment with 8-methoxypsoralen and UVA irradiation
YDR235W/PRP42	U1 snRNP protein involved in splicing, required for U1 snRNP biogenesis; contains multiple tetriatricopeptide repeats
YER151C/UBP3	Ubiquitin-specific protease that interacts with Bre5p to co-regulate anterograde and retrograde transport between the ER and Golgi; inhibitor of gene silencing; cleaves ubiquitin fusions but not polyubiquitin; also has mRNA binding activity
YMR275C/BUL1	Ubiquitin-binding component of the Rsp5p E3-ubiquitin ligase complex, functional homolog of Bul2p, disruption causes temperature-sensitive growth, overexpression causes missorting of amino acid permeases
YKL114C/APN1	Major apurinic/apyrimidinic endonuclease, 3'-repair diesterase involved in repair of DNA damage by oxidation and alkylating agents; also functions as a 3'-5' exonuclease to repair 7,8-dihydro-8-oxodeoxyguanosine
YOL081W/IRA2	GTPase-activating protein that negatively regulates RAS by converting it from the GTP- to the GDP-bound inactive form, required for reducing cAMP levels under nutrient limiting conditions, has similarity to Ira1p and human neurofibromin
YPR049C/ATG11	Adapter protein for pexophagy and the cytoplasm-to-vacuole targeting (Cvt) pathway; directs receptor-bound cargo to the phagophore assembly site (PAS) for packaging into vesicles; required for recruiting other proteins to the (PAS)
YGL095C/VPS45	Protein of the Sec1p/Munc-18 family, essential for vacuolar protein sorting; required for the function of Pep12p and the early endosome/late Golgi SNARE Tlg2p; essential for fusion of Golgi-derived vesicles with the prevacuolar compartment
YDR456W/NHX1	Na+/H+ and K+/H+ exchanger, required for intracellular sequestration of Na+ and K+; located in the vacuole and late

	endosome compartments; required for osmotolerance to acute hypertonic shock and for vacuolar fusion
YKL197C/PEX1	AAA-peroxin that heterodimerizes with AAA-peroxin Pex6p and participates in the recycling of peroxisomal signal receptor Pex5p from the peroxisomal membrane to the cystosol; induced by oleic acid and upregulated during anaerobiosis
YIL068C/SEC6	Essential 88kDa subunit of the exocyst complex, which mediates polarized targeting of secretory vesicles to active sites of exocytosis; dimeric form of Sec6p interacts with Sec9p in vitro and inhibits t-SNARE assembly
YOR326W/MYO2	One of two type V myosin motors (along with MYO4) involved in actin-based transport of cargos; required for the polarized delivery of secretory vesicles, the vacuole, late Golgi elements, peroxisomes, and the mitotic spindle
YNR045W/PET494	Mitochondrial translational activator specific for the COX3 mRNA, acts together with Pet54p and Pet122p; located in the mitochondrial inner membrane
YJR107W/LIH1	Putative protein of unknown function; has sequence or structural similarity to lipases
YPL268W/PLC1	Phospholipase C, hydrolyzes phosphatidylinositol 4,5-biphosphate (PIP2) to generate the signaling molecules inositol 1,4,5-triphosphate (IP3) and 1,2-diacylglycerol (DAG); involved in regulating many cellular processes
YJL062W/LAS21	Integral plasma membrane protein involved in the synthesis of the glycosylphosphatidylinositol (GPI) core structure; mutations affect cell wall integrity
YCR042C/TAF2	TFIID subunit (150 kDa), involved in RNA polymerase II transcription initiation

Supplementary Table 5. Annotations of top hit loci from barcoded RH-seq of thermotolerance. Shown are hits from thermotolerance mapping by barcoded RH-seq (Table S7) that met quality control thresholds and at which disruption of the *S. cerevisiae* allele compromised thermotolerance to a greater extent than did disruption of the *S. paradoxus* allele in the interspecific hybrid.

gene	D_s	D_n	P_s	P_n	NI	*p*	adjusted *p*
YBR136W/MEC1	293	38	770	363	3.63	5.71×10^{-15}	1.69×10^{-12}
YLR422W/DCK1	122	26	335	280	3.92	1.13×10^{-10}	8.70×10^{-9}
YOR326W/MYO2	35	0	1486	1273	inf	5.15×10^{-10}	2.77×10^{-3}
YAL026C/DRS2	32	1	1399	1595	36.48	5.81×10^{-10}	2.97×10^{-8}
YMR207C/HFA1	77	16	2215	2103	4.57	6.16×10^{-10}	3.08×10^{-8}
YOR371C/GPB1	79	18	347	323	4.09	1.42×10^{-8}	4.02×10^{-7}
YML099C/ARG81	96	16	351	247	4.22	1.61×10^{-8}	4.43×10^{-7}
YJL062W/LAS21	74	10	170	131	5.70	2.19×10^{-8}	5.77×10^{-7}
YPL268W/PLC1	112	16	245	146	4.17	4.57×10^{-8}	1.05×10^{-6}
YCR042C/TAF2	155	45	661	470	2.45	1.70×10^{-7}	2.90×10^{-6}
YPL254W/HFI1	76	10	150	106	5.37	1.66×10^{-7}	2.90×10^{-6}
YPR049C/ATG11	131	45	448	395	2.57	1.67×10^{-7}	2.90×10^{-6}
YNL049C/SFB2	32	1	773	587	24.30	3.22×10^{-7}	4.88×10^{-6}
YKL114C/APN1	44	3	120	87	10.63	9.80×10^{-7}	1.19×10^{-5}
YIL068C/SEC6	80	5	271	111	6.55	1.36×10^{-6}	1.52×10^{-5}
YGR198W/YPP1	79	25	295	282	3.02	1.97×10^{-6}	2.10×10^{-5}
YDR375C/BCS1	67	4	107	54	8.45	2.20×10^{-6}	2.28×10^{-5}
YKL017C/HCS1	78	13	219	143	3.92	3.56×10^{-6}	3.28×10^{-5}
YDR235W/PRP42	61	7	177	109	5.37	5.34×10^{-6}	4.57×10^{-5}
YDR180W/SCC2	173	72	523	431	1.98	6.47×10^{-6}	5.35×10^{-5}
YMR167W/MLH1	94	26	286	209	2.64	2.47×10^{-5}	1.54×10^{-4}
YKL197C/PEX1	132	58	415	351	1.92	1.55×10^{-4}	6.57×10^{-4}
YMR078C/CTF18	73	26	295	260	2.47	1.63×10^{-4}	6.82×10^{-4}
YGL095C/VPS45	67	14	202	122	2.89	3.72×10^{-4}	1.30×10^{-3}
YMR094W/CTF13	54	21	152	158	2.67	4.44×10^{-4}	1.50×10^{-3}
YNL132W/KRE33	75	4	187	52	5.21	5.13×10^{-4}	1.69×10^{-3}
YDR103W/STE5	93	56	344	373	1.80	1.57×10^{-3}	4.16×10^{-3}
YOR092W/ECM3	16	1	648	498	12.30	1.92×10^{-3}	4.86×10^{-3}
YNR045W/PET494	61	20	177	136	2.34	2.15×10^{-3}	5.32×10^{-3}
YJR107W/LIH1	19	1	102	61	11.36	2.40×10^{-3}	5.79×10^{-3}
YLR397C/AFG2	102	27	288	155	2.03	2.56×10^{-3}	6.10×10^{-3}
YGL082W/MIY1	41	8	132	84	3.26	2.60×10^{-3}	6.17×10^{-3}
YOR091W/TMA46	40	6	103	46	2.98	0.0212	0.0351
YDR456W/NHX1	87	13	187	59	2.11	0.0278	0.0443
YDR508C/GNP1	3	0	959	797	inf	0.2562	0.3009
YIL152W/VPR1	19	11	70	59	1.46	0.4184	0.4669

Supplementary Table 6. Whole-gene tests for evidence of non-neutral protein evolution at thermotolerance loci. Each row reports results from the McDonald-Kreitman test on sequences from strains of European populations of *S. cerevisiae* and *S. paradoxus* of the indicated top hit from barcoded RH-seq mapping of thermotolerance. D_s, number of sites of synonymous nucleotide divergence between species; number of sites of D_n, nonsynonymous nucleotide divergence between species; P_s, number of sites of synonymous nucleotide polymorphisms within species; P_n, number of sites of nonsynonymous nucleotide polymorphisms within species. NI, neutrality index. The sixth column reports the *p*-value from a Fisher's exact test on D_s, D_n, P_s, and P_n, and the seventh column reports the adjusted *p*-value after applying the Benjamini-Hochberg correction for multiple hypothesis testing. All loci exhibited NI > 1, corresponding to a dearth of divergent amino acid changes relative to synonymous changes and polymorphisms.

Supplementary note: optimizing the barcoded yeast piggyBac system

With the goal of maximizing transposition efficiency in barcoded yeast piggyBac, we optimized two facets of the system by making changes to the piggyBac vector (Figure S2A; all cloning by Genscript, Inc.) and assaying transposition as follows.

We reasoned that we could optimize transposition efficiency independent of barcode location by making codon and/or amino acid changes to the transposase coding region. For this purpose, in the un-barcoded, BbsI-ready piggyBac plasmid pCW328 (Figure S2A), we codon-optimized the transposase coding region for expression in S. cerevisiae (using a proprietary method from Genscript), yielding the test plasmid pJC10 (Figure S2A, top right). We then used the latter plasmid as a backbone for the introduction of amino acid changes to the transposon coding region (I30V, S130P, S165S, M282V, S509G, N538K, N571S) which (Yusa et al. 2011) reported to result in a hyperactive piggyBac in mammalian cells, resulting in the test plasmid pJC11 (Figure S2A, center right). For these two plasmids, we assessed transposition efficiency with test transformations as in Methods, but at one tenth the scale, using only one tenth of the cell pellet of each 50 mL log-phase hybrid culture at OD 0.9. Approximately 100 individual 5FOA+ colonies were patched onto YPD+ G418 300ug/ml plates and incubated overnight at 28oC. Transposition efficiency was calculated as the proportion of G418 patches that grew on 5FOA media out of the total patches on the G418 plate. Results revealed no evidence for improvement in transposition by either manipulation (Figure S2B).

Separately, to explore the potential position of barcodes in piggyBac, we inserted a test barcode in the transposon of the un-barcoded, BbsI-ready piggyBac plasmid pCW328 (Figure S2A) in each of two locations. In one scheme, a 64-nucleotide segment, containing a single barcode flanked by universal priming regions and custom two-nucleotide overhang sequences, was inserted between the 3’ end of the left arm of the transposon and the 5’ end of the TEF promoter of the kanamycin resistance cassette, nucleotides 353-354 of the pCW328 vector, resulting in pJC4 (Figure S2A, bottom left). This position of insertion was chosen to avoid the two internal binding sites in the left arm of piggyBac (Morellet et al. 2018). In a second scheme, the 64-nucleotide segment replaced 64 endogenous nucleotides inside the end of the right arm of the transposon (nucleotides 1984-2047 of the pCW328 vector), resulting in pJC9 (Figure S2A, bottom right). The latter was chosen to avoid compromising the region which (Morellet et al. 2018) suggest functions as a C-terminal DNA-binding domain in the transposon based on DNAse I footprinting data, and which has previously been shown to constitute a minimal transposable element (Mitra et al. 2008; Meir et al. 2011; Solodushko et al. 2014). Transposition assays results revealed that the piggyBac bearing the replacement in the right arm of the transposon, pJC9, performed best (Figure S2B), and we used this scheme for the final barcoding of piggyBac for RH-seq at production scale as detailed in Methods.

Chapter 4

Temperature-dependent genetics of thermotolerance between yeast species

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

Abrams, M.B., and Brem, R.B. Preprint. Temperature-dependent genetics of thermotolerance between yeast species. https://doi.org/10.1101/2022.01.11.475859.

ABSTRACT

Many traits of industrial and basic biological interest arose long ago, and manifest now as fixed differences between a focal species and its reproductively isolated relatives. In these systems, extant individuals can hold clues to the mechanisms by which phenotypes evolved in their ancestors. We harnessed yeast thermotolerance as a test case for such molecular-genetic inferences. In viability experiments, we showed that extant *Saccharomyces* cerevisiae survived at temperatures where cultures of its sister species *S. paradoxus* died out. Then, focusing on loci that contribute to this difference, we found that the genetic mechanisms of high-temperature growth changed with temperature. We also uncovered an enrichment of low-frequency variants at thermotolerance loci in *S. cerevisiae* population sequences, suggestive of a history of non-neutral forces acting at these genes. We interpret our results in light of a model of gradual acquisition of thermotolerance in the *S. cerevisiae* lineage by positive selection along a temperature cline. We propose that in an ancestral *S. cerevisiae* population, alleles conferring defects at a given temperature would have been resolved by adaptive mutations, expanding the range and setting the stage for further temperature advances. Together, our results and interpretation underscore the power of genetic approaches to explore how an ancient trait came to be.

INTRODUCTION

A central goal of evolutionary genetics is to understand how nature builds new phenotypes. Thanks to advances in statistical genetics and experimental evolution, mechanisms of trait evolution over relatively short timescales have come well within reach in the modern literature. By contrast, longer-term innovations have posed a more profound challenge for the field (Orr, 2001). In principle, for a phenotype that originated long ago and manifests now as a fixed difference between species, evolution could have refined the character along the entire divergence time of the respective taxa. In landmark cases, candidate-gene studies have shed light on suites of mutational changes of this kind between species, at a given model locus. This includes the order by which adaptive alleles were likely acquired, and/or the functional pressures that drove them (Anderson et al., 2015; Baldwin et al., 2014; Bridgham et al., 2009;

Daugherty et al., 2016; Dong et al., 2014; Liu et al., 2018; Pillai et al., 2020; Prieto-Godino et al., 2021; Sayou et al., 2014; Siddiq and Thornton, 2019; Starr et al., 2018; Sulak et al., 2016; Xie et al., 2018). But, in most cases, any factor pursued by such a candidate-gene approach only represents part of the complex genetic architecture of an ancient trait. We still know relatively little about how evolution coordinates multiple adaptive loci over deep divergences.

In the search for evolutionary principles, genetically tractable model systems can be of great utility. *Saccharomyces* yeasts are well suited for this purpose, and environmental yeast isolates have been studied extensively for their innovations within and between species (Hittinger, 2013). Thermotolerance is of particular interest because it tracks with phylogeny across the 20 million years of the *Saccharomyces* radiation (Gonçalves et al., 2011; Salvadó et al., 2011). Even the two most recent branches of the phylogeny exhibit a robust difference in this phenotype: *S. cerevisiae* acquired the ability to grow at temperatures near 40°C in the five million years since it diverged from its sister species, *S. paradoxus* (Gonçalves et al., 2011; Salvadó et al., 2011; Sweeney et al., 2004; Williams et al., 2015). Previously, a whole-genome mapping scheme was used to identify housekeeping genes at which variation between *S. cerevisiae* and *S. paradoxus* impacts growth at the high end of the *S. cerevisiae* temperature range (Weiss et al., 2018b). These loci exhibit striking sequence differences between the species, and conservation in *S. cerevisiae*, consistent with a history of positive selection on pro-thermotolerance alleles (Abrams et al., 2021b, 2021a; Weiss et al., 2018b). However, we have as yet little insight into the ecological dynamics by which this model trait evolved along the *S. cerevisiae* lineage.

Cases of adaptation across temperature clines are a mainstay of the evolutionary genetics literature (Calfee et al., 2021; Dudaniec et al., 2018; Endler, 2020; Key et al., 2018; Machado et al., 2021; Mimura et al., 2013; Prasad et al., 2011; Robin et al., 2017; Savolainen et al., 2013; Tepolt and Palumbi, 2020; Turner et al., 2008). Here we sought to explore the relevance of such a mechanism to the events by which *S. cerevisiae* gained its maximal thermotolerance phenotype. Given that we have no access to genotypes representing ancient intermediates between this species and *S. paradoxus*, we designed a strategy to interrogate the genetics of extant strains, focusing on contributing genes of major effect. We surveyed gene-environment interactions by these thermotolerance loci across warm temperatures, complementing previous studies of interspecies variation at a single high temperature (Abrams et al., 2021a; Weiss et al., 2018b). And we investigated the frequency of variants at these loci with a population-genomic approach.

MATERIALS AND METHODS

Dose response growth assay

For growth measurements in Figure 2, we assayed *S. paradoxus* Z1, S. cerevisiae DBVPG1373, and *S. cerevisiae* DBVPG1373 with the *S. paradoxus* Z1 allele of *ESP1*, *MYO1*, *AFG2*, or *CEP3* swapped in at the endogenous locus from (Weiss et al., 2018b) (Table S2) as follows. Each strain was streaked from a -80°C freezer stock onto a yeast peptone dextrose (YPD) agar plate and incubated at room temperature for 3 days. For each biological replicate, a single colony was inoculated into 5 mL liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate pre-cultures. Each pre-culture was backdiluted into YPD at an OD_{600}/mL of 0.05 and grown for an additional 5.5-6 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. Each pre-culture was again back-diluted into 10 mL YPD in 1-inch diameter glass tubes with a target OD_{600}/mL of 0.05; the actual OD_{600}/mL of each was measured, after which it was grown at a temperature of interest (28°C or 35-38°C) with shaking at 200rpm for 24 hours, and OD_{600}/mL was measured again. The growth efficiency for each replicate was calculated as the difference between these final and initial OD_{600}/mL values. We used the growth efficiency from all days and all temperatures of a given strain s as input into a two-factor type 2 ANOVA test for a temperature-by-strain effect comparing s with *S. cerevisiae*.

For growth measurements of *ESP1* swap strains with different donors, as reported in Figure S2, cultures were grown and measured as above, except that the only temperature was 36°C.

Viability Assay

For the survey of viability phenotypes at high temperatures across wild-type isolates in Figure 2, strains were streaked out and a colony of each was pre-cultured in liquid as for 39°C growth above, except that the initial pre-culture to achieve saturation lasted 48 hours. Each pre-culture was back-diluted into 10 mL of YPD to reach an OD_{600}/mL of 0.05 and then cultured for 24 hours at the temperature of interest (28°C - 39°C). The viability of both the precultures and the cultures after 24 hours at the temperature of interest were measured with a spotting assay, where we diluted aliquots from the culture in a 1:10 series and spotting 3 µL of each dilution for growth on a solid YPD plate. After incubation at 28°C for two days, we used the dilution corresponding to the densest spot that was not a lawn for to determine viability: we counted the number of colonies in each of the two technical replicate spots, formulated the number of colony forming units per mL of undiluted culture (CFU/mL). We determined the change in viable cells by subtracting the number of cells in the culture at the initial time point from that at the final timepoint, based on the CFU/mL count and the culture volume. We evaluated the significance of the difference between *S. cerevisiae* and *S. paradoxus* at a given temperature using a one-sided Mann-Whitney U test.

Tajima's D in Wine/European *S. cerevisiae*

Tajima's D tabulates the difference between the average number of differences in pairs of sequences in a population sample and the number of variant sites in the sample. When the latter is of much bigger magnitude and D is negative, it indicates that variation in the sample is accounted for mostly by rare alleles. This pattern is expected some time after a selective sweep when de novo mutations arise on the swept haplotype; it can also reflect weak purifying selection or a history of population expansion (Suzuki, 2010). For instance, the Wine/European population from (Peter et al., 2018) which we focus on in this study has a low median Tajima's D, a phenomenon which the authors of that work interpret in the context of a model of population expansion.

We used a resampling test to assess enrichment trends in Tajima's D in genes of interest against a gnomic null. This scheme normalizes out impacts on Tajima's D values from events which affect the whole genome, such as a population expansion; that said, formally, any results from such an empirical outlier-based analysis serve as suggestive rather than conclusive evidence for non-neutral evolution (Teshima et al., 2006; Thornton and Jensen, 2007). We calculated Tajima's D for each gene from the coding start to the coding stop of each gene, using code drawn from used VCF-kit (Cook and Andersen, 2017) on an input of the VCFs reporting inheritance in the strains of a given *S. cerevisiae* population from (Peter et al., 2018). For the resampling test, we sampled 10,000 random cohorts of genes from the genome with the same number of essential and nonessential genes as our thermotolerance cohort (Winzeler et al., 1999), and we used as an empirical P-value the proportion of random cohorts where the median Tajima's D was less than or equal to that of our thermotolerance cohort. Application of this test to the Wine/European *S. cerevisiae* population (362 strains) is reported in Figure 3 of the main text. Applied to the Mosaic Region 3 (113 strains), Mixed Origin (72 strains), Sake (47 strains), and Brazilian Bioethanol (35 strains), the four next most deeply sampled populations from (Peter et al., 2018), this test for enrichment of low Tajima's D among our four focal thermotolerance genes yielded P = 0.8351, 0.975, 0.8624, and 0.0348 respectively.

RESULTS

For an initial study of the genetics of yeast species variation in temperature response, we chose to harness DBVPG1373, an *S. cerevisiae* isolate from Dutch soil, and Z1, an *S. paradoxus* isolate from an oak tree in England. We anticipated that detailed analyses using these strains, as representatives of their respective species, could accelerate the discovery of more general principles (Abrams et al., 2021b, 2021a; Weiss et al., 2018b). We developed an assay quantifying cell viability in a given liquid culture before and after incubation at a temperature of interest, and we implemented this approach for each species in turn. The results revealed an advantage for *S. cerevisiae* over *S. paradoxus* at temperatures above 35°C (Figure 1), consistent with previous growth-based surveys (Gonçalves et al., 2011; Salvadó et al., 2011; Sweeney et al., 2004; Williams et al., 2015). *S. cerevisiae* maintained viability at all temperatures tested, whereas by 37°C, *S. paradoxus* became inviable, with no evidence for spontaneous rescue even over long incubation times (Figure 1). Using the latter as a window onto the phenotype of the

common ancestor of *S. cerevisiae* and *S. paradoxus*, we would envision that the latter ancient population would have gone extinct rather than adapting, if exposed to temperatures at the high end of the range tolerated by modern *S. cerevisiae*.

We next aimed to investigate the genetics of temperature response as it differs between extant *S. cerevisiae* and *S. paradoxus*, to motivate inferences about the evolution of the maximal thermotolerance trait. We focused on four genes—the cell division factors *ESP1*, *MYO1*, and *CEP3*, and the ribosome maturation factor *AFG2*—where alleles from modern *S. paradoxus* compromise growth at 39°C (Weiss et al., 2018b). We made use of strains of the *S. cerevisiae* DBVPG1373 background harboring the allele of each gene in turn from *S. paradoxus* Z1. In each, we measured the growth phenotype as a dose-response across temperatures, and observed a drop as temperature increased (Figure 2). In the *S. cerevisiae* background, *S. paradoxus* Z1 alleles eroded growth efficiency at temperatures well below the hard limit of viability for the Z1 wild-type (~38°C), an effect that reached significance for three of our four focal genes (Figure 2). We conclude that many problems posed by *S. paradoxus* Z1 alleles at the high end of our temperature range also manifest to a lesser extent at lower temperatures. Inspecting the shape of the temperature dose-responses of allelic effects, we noted quantitative differences between our loci. At the chromatid separase gene *ESP1*, the Z1 allele conferred an appreciable drop in growth efficiency at 36°C and supported almost no growth by 37°C (Figure 2B). By contrast, at *AFG2*, the *S. paradoxus* Z1 allele was sufficient for growth approximating that of *S. cerevisiae* until 38°C (Figure 2D). The dose-response of allelic effects at *MYO1*, encoding a class II myosin heavy chain, lay between these two extremes (Figure 2C). This differential susceptibility to temperature across the genes of our set likely reflects distinguishing properties of their structure and function, and of the interspecies variants they harbor.

We reasoned that trends from our temperature dose-response approach would be most informative when they were conserved across *S. paradoxus* as a species. To pursue this, we earmarked *ESP1*, whose *S. paradoxus* Z1 allele had exhibited the sharpest falloff with temperature among the genes of our set (Figure 2B). We repeated our growth efficiency experiments in strains of *S. cerevisiae* DBVPG1373 harboring *ESP1* from other *S. paradoxus* donors beside the Z1 strain. These transgenics, which phenocopy wild-type DBVPG1373 at 28°C (Weiss et al., 2018b), all dropped off in growth efficiency by 36°C (Figure S2), as expected if the temperature preference of *ESP1* were ancestral to, and shared across, extant *S. paradoxus* populations. Together, these dose-response results make clear that the functions of *S. paradoxus* alleles at our focal genes break down at distinct temperatures between 35°C and 38°C—suggesting similar gene-by-environment effects in the ancestor of *S. cerevisiae* and *S. paradoxus*, if it sampled a range of temperature conditions over evolutionary history.

To gain insight into past and current selective pressures on thermotolerance loci, we turned to a molecular-evolution approach. Previous work on *Saccharomyces* thermotolerance genes has emphasized sequence divergence between species (Abrams et al., 2021b, 2021a; Weiss et al., 2018b). For a complementary focus on population variation within *S. cerevisiae*, we analyzed the Tajima's D statistic, which formulates properties of sequence variants in a population into a test of how well the sequence fits the expectations under a neutral evolutionary model (Biswas and Akey, 2006; Tajima, 1989). We first developed a resampling-based scheme that compares Tajima's D between a gene cohort of interest and a genomic null. This enables a non-parametric assessment of significance and normalizes out potential effects of demography on the statistic (see Methods). Expecting that our test would have maximal power in a data set of large sample size, we focused on the most deeply-sampled *S. cerevisiae* population in current compendia, comprising isolates from vineyards and European soil (Peter et al., 2018). Examining our four focal thermotolerance genes, we detected an enrichment for low, negative Tajima's D at these loci across *S. cerevisiae* genomes (Figure 3 and Table S1), reporting an excess of rare variants—as expected after a selective sweep, or under constraints from purifying selection (Biswas and Akey, 2006; Suzuki, 2010). We repeated this analysis using more comprehensive sets of hits from interspecies thermotolerance screens (Abrams et al., 2021b, 2021a; Weiss et al., 2018b), and detected strong signal for low, negative Tajima's D at these loci in vineyard/European *S. cerevisiae* in every case (Table S1). Interestingly, *ESP1* exhibited the most negative Tajima's D value among all thermotolerance genes (Table S1), dovetailing with the strong effect of variation at this gene in phenotypic analyses (Weiss et al. 2018 and Figure 2). The trend for low, negative Tajima's D in thermotolerance genes was detectable but not consistent across other shallowly-sampled populations of *S. cerevisiae* (see Methods), potentially reflecting weaker power or weaker selection in the latter relative to vineyard/European strains. In either case, our data establish that thermotolerance gene variation in some modern *S. cerevisiae* populations is consistent with a history of non-neutral evolution.

DISCUSSION

Most trait differences that define long-diverged species are likely the product of suites of unlinked variants that have come together over long timescales. Tracing the ancient evolutionary events at such loci remains a central challenge in the field. In this work, we have characterized thermotolerance genes in extant yeast strains and species to inform models of the evolution of the trait in *S. cerevisiae*.

Our data have shown that temperatures below the high end of the *S. cerevisiae* range are lethal for *S. paradoxus*. This complements previous surveys of the *Saccharomyces* clade using growth-based assays (Gonçalves et al., 2011; Salvadó et al., 2011; Sweeney et al., 2004) and reveals that, under high temperature conditions, a given culture of *S. paradoxus* will die off rather than adapting. Assuming similar behavior in an

ancestor of *S. cerevisiae* before the rise of its modern thermotolerance profile, we infer that the latter event was unlikely to be precipitated by a single jump to a hot growth environment, long ago in history. Rather, we favor the hypothesis that thermotolerance evolution through the *S. cerevisiae* lineage proceeded along a temperature cline, as has been documented in elegant local adaptation case studies (Dudaniec et al., 2018; Key et al., 2018; Mimura et al., 2013; Robin et al., 2017; Tepolt and Palumbi, 2020). In this scenario, a well-adapted ancestral *S. cerevisiae* population in a given temperature niche would have acquired variants that resolved defects manifesting in slightly warmer conditions, and expanded its range. As complementary possibility, *S. cerevisiae* in its original geographic location could have acquired variants that resolved defects manifesting as local temperature gradually increased. If the *S. cerevisiae* lineage did go through a series of such transitions in acquiring maximal thermotolerance, it would echo the principle that gradual exposure to increasing stresses fosters adaptation and reduces the risk of extinction, relative to a sudden, high dose of stress (Lindsey et al., 2013).

We have also leveraged results from genetic mapping of *S. cerevisiae* thermotolerance to trace the temperature dependence of contributing genes. This contrasts with approaches that search genomes for variants associated with environment or population variables (Hoban et al., 2016; Rellstab et al., 2015), in that we focus directly on bona fide causal determinants of the trait of interest. Our strategy has allowed us to discern differences between thermotolerance loci in terms of the temperatures at which alleles from modern *S. paradoxus* fail in their growth functions. In other words, the genetic mechanisms of growth change with temperature in this system. If such effects were at play in the *S. cerevisiae* ancestor, as it sampled conditions and niches during the acquisition of thermotolerance, each locus would have come under selective pressure at the temperature where its defect manifested. This phenomenon, by which the weakest point in the genetic network targeted by evolution changes as conditions change, has been termed “whack-a-mole” dynamics (Shin and MacCarthy, 2015). Ultimately, a complete model of adaptation across time and environment will take into account the shifts in the evolutionary landscape from this effect as well as from constant changes in genetic background (Starr and Thornton, 2016).

If the ancient *S. cerevisiae* population did adapt progressively along a temperature cline, what would the ecology have been? In principle, migrants from temperate physical locales could have advanced to warmer and warmer locales, perhaps terminating in the hot East Asian environments to which the ancestor of modern *S. cerevisiae* has been traced (Peter et al., 2018). Alternatively, the trait syndrome in this species—a unique ability to tolerate ethanol as well as heat, with both given off by fermentative metabolism—could have arisen as a specialization to kill off microbial competitors in nutrient-rich substrates, regardless of the endemic temperature of the location (Goddard, 2008; Salvadó et al., 2011). If so, variants would have been acquired, potentially over millions of generations, gradually to ratchet up fermentative activity and

tolerance to its byproducts, with our genetic insights thus far largely restricted to the latter.

Our work also leaves open the dating of any such events. In several *S. cerevisiae* populations, we have uncovered an enrichment of rare variants at thermotolerance genes. This signature of non-neutral evolution can be interpreted at a given locus as evidence for a relatively recent sweep of a positively selected haplotype, or for negative selection to maintain a fitness-relevant haplotype that arose long ago. We favor the latter hypothesis, given that prior work across populations has also made clear that *S. cerevisiae* alleles at our focal genes are partly sufficient for maximal thermotolerance, conserved within the species, and divergent from *S. paradoxus* (Abrams et al., 2021a; Weiss et al., 2018b). We thus propose that many thermotolerance alleles were acquired in ancient selective sweeps, before the emergence of modern *S. cerevisiae,* and have been maintained since then by purifying selection. That said, later refinements in particular populations of *S. cerevisiae* may also have strengthened beneficial facets of the trait, added regulatory tuning, or eliminated antagonistic pleiotropic "side effects" that were niche-specific. A comprehensive genetic and ecological reconstruction of this history may be out of our current grasp, especially in light of the caveats of our approach using a laboratory setting and extant strain backgrounds. Nonetheless, our data add compelling detail to an emerging consensus view of how evolution built maximal thermotolerance in *S. cerevisiae*.

ACKNOWLEDGEMENTS

This work was supported by NSF GRFP DGE 1752814 to M.A. and NIH R01 GM120430 to R.B.B. The authors thank Faisal AlZaben, Abel Duarte, and Jude Edwards for experimental support and Adam Arkin for his generosity with computational resources.

REFERENCES

Abrams, M.B., Chuong, J.N., AlZaben, F., Dubin, C.A., Skerker, J.M., and Brem, R.B. (2021a). Barcoded reciprocal hemizygosity analysis via sequencing illuminates the complex genetic basis of yeast thermotolerance.

Abrams, M.B., Dubin, C.A., AlZaben, F., Bravo, J., Joubert, P.M., Weiss, C.V., and Brem, R.B. (2021b). Population and comparative genetics of thermotolerance divergence between yeast species. G3 Genes|Genomes|Genetics.

Anderson, D.W., McKeown, A.N., and Thornton, J.W. (2015). Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. ELife 4, e07864.

Baldwin, M.W., Toda, Y., Nakagita, T., O'Connell, M.J., Klasing, K.C., Misaka, T., Edwards, S.V., and Liberles, S.D. (2014). Sensory biology. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. Science 345, 929–933.

Biswas, S., and Akey, J.M. (2006). Genomic insights into positive selection. Trends in Genetics 22, 437–446.
Bridgham, J.T., Ortlund, E.A., and Thornton, J.W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461, 515–519.
Calfee, E., Gates, D., Lorant, A., Perkins, M.T., Coop, G., and Ross-Ibarra, J. (2021). Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. PLOS Genetics 17, e1009810.
Cook, D.E., and Andersen, E.C. (2017). VCF-kit: assorted utilities for the variant call format. Bioinformatics 33, 1581–1582.
Daugherty, M.D., Schaller, A.M., Geballe, A.P., and Malik, H.S. (2016). Evolution-guided functional analyses reveal diverse antiviral specificities encoded by IFIT1 genes in mammals. ELife 5, e14228.
Dong, S., Stam, R., Cano, L.M., Song, J., Sklenar, J., Yoshida, K., Bozkurt, T.O., Oliva, R., Liu, Z., Tian, M., et al. (2014). Effector specialization in a lineage of the Irish potato famine pathogen. Science 343, 552–555.
Dudaniec, R.Y., Yong, C.J., Lancaster, L.T., Svensson, E.I., and Hansson, B. (2018). Signatures of local adaptation along environmental gradients in a range-expanding damselfly (Ischnura elegans). Molecular Ecology 27, 2576–2593.
Endler, J.A. (2020). Geographic Variation, Speciation and Clines. (MPB-10), Volume 10 (Princeton University Press).
Goddard, M.R. (2008). Quantifying the complexities of Saccharomyces cerevisiae's ecosystem engineering via fermentation. Ecology 89, 2077–2082.
Gonçalves, P., Valério, E., Correia, C., Almeida, J.M.G.C.F. de, and Sampaio, J.P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric Saccharomyces Species. PLOS ONE 6, e20739.
Hittinger, C.T. (2013). Saccharomyces diversity and evolution: a budding model genus. Trends Genet 29, 309–317.
Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L., Reed, L.K., Storfer, A., and Whitlock, M.C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am Nat 188, 379–397.
Key, F.M., Abdul-Aziz, M.A., Mundry, R., Peter, B.M., Sekar, A., D'Amato, M., Dennis, M.Y., Schmidt, J.M., and Andrés, A.M. (2018). Human local adaptation of the TRPM8 cold receptor along a latitudinal cline. PLOS Genetics 14, e1007298.
Lindsey, H.A., Gallie, J., Taylor, S., and Kerr, B. (2013). Evolutionary rescue from extinction is contingent on a lower rate of environmental change. Nature 494, 463–467.
Liu, Q., Onal, P., Datta, R.R., Rogers, J.M., Schmidt-Ott, U., Bulyk, M.L., Small, S., and Thornton, J.W. (2018). Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. Elife 7.
Machado, H.E., Bergland, A.O., Taylor, R., Tilk, S., Behrman, E., Dyer, K., Fabian, D.K., Flatt, T., González, J., Karasov, T.L., et al. (2021). Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in Drosophila. ELife 10, e67577.

Mimura, M., Ono, K., Goka, K., and Hara, T. (2013). Standing variation boosted by multiple sources of introduction contributes to the success of the introduced species, Lotus corniculatus. Biol Invasions 15, 2743–2754.
Orr, H.A. (2001). The genetics of species differences. Trends Ecol Evol 16, 343–350.
Peter, J., Chiara, M.D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344.
Pillai, A.S., Chandler, S.A., Liu, Y., Signore, A.V., Cortez-Romero, C.R., Benesch, J.L.P., Laganowsky, A., Storz, J.F., Hochberg, G.K.A., and Thornton, J.W. (2020). Origin of complexity in haemoglobin evolution. Nature 581, 480–485.
Prasad, A., Croydon-Sugarman, M.J.F., Murray, R.L., and Cutter, A.D. (2011). Temperature-Dependent Fecundity Associates with Latitude in Caenorhabditis Briggsae. Evolution 65, 52–63.
Prieto-Godino, L.L., Schmidt, H.R., and Benton, R. (2021). Molecular reconstruction of recurrent evolutionary switching in olfactory receptor specificity. ELife 10, e69732.
Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M., and Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. Mol Ecol 24, 4348–4370.
Robin, C., Andanson, A., Saint-Jean, G., Fabreguettes, O., and Dutech, C. (2017). What was old is new again: thermal adaptation within clonal lineages during range expansion in a fungal pathogen. Molecular Ecology 26, 1952–1963.
Salvadó, Z., Arroyo-López, F.N., Guillamón, J.M., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus Saccharomyces. Appl Environ Microbiol 77, 2292–2302.
Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics 14, 807–820.
Sayou, C., Monniaux, M., Nanao, M.H., Moyroud, E., Brockington, S.F., Thévenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., et al. (2014). A Promiscuous Intermediate Underlies the Evolution of LEAFY DNA Binding Specificity. Science 343, 645–648.
Shin, J., and MacCarthy, T. (2015). Antagonistic Coevolution Drives Whack-a-Mole Sensitivity in Gene Regulatory Networks. PLoS Comput Biol 11, e1004432.
Siddiq, M.A., and Thornton, J.W. (2019). Fitness effects but no temperature-mediated balancing selection at the polymorphic Adh gene of Drosophila melanogaster. PNAS 116, 21634–21640.
Starr, T.N., and Thornton, J.W. (2016). Epistasis in protein evolution. Protein Science 25, 1204–1218.
Starr, T.N., Flynn, J.M., Mishra, P., Bolon, D.N.A., and Thornton, J.W. (2018). Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. PNAS 115, 4453–4458.
Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D., and Lynch, V.J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. ELife 5, e11994.

Suzuki, Y. (2010). Statistical methods for detecting natural selection from genomic data. Genes Genet Syst 85, 359–376.
Sweeney, J.Y., Kuehne, H.A., and Sniegowski, P.D. (2004). Sympatric natural Saccharomyces cerevisiae and S. paradoxus populations have different thermal growth profiles. FEMS Yeast Res. 4, 521–525.
Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 123, 585–595.
Tepolt, C.K., and Palumbi, S.R. (2020). Rapid Adaptation to Temperature via a Potential Genomic Island of Divergence in the Invasive Green Crab, Carcinus maenas. Frontiers in Ecology and Evolution 8, 411.
Teshima, K.M., Coop, G., and Przeworski, M. (2006). How reliable are empirical genomic scans for selective sweeps? Genome Res 16, 702–712.
Thornton, K.R., and Jensen, J.D. (2007). Controlling the False-Positive Rate in Multilocus Genome Scans for Selection. Genetics 175, 737–750.
Turner, T.L., Levine, M.T., Eckert, M.L., and Begun, D.J. (2008). Genomic Analysis of Adaptive Differentiation in Drosophila melanogaster. Genetics 179, 455–473.
Weiss, C.V., Roop, J.I., Hackley, R.K., Chuong, J.N., Grigoriev, I.V., Arkin, A.P., Skerker, J.M., and Brem, R.B. (2018). Genetic dissection of interspecific differences in yeast thermotolerance. Nat Genet 50, 1501–1504.
Williams, K.M., Liu, P., and Fay, J.C. (2015). Evolution of ecological dominance of yeast species in high-sugar environments. Evolution 69, 2079–2093.
Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science 285, 901–906.
Xie, J., Li, Y., Shen, X., Goh, G., Zhu, Y., Cui, J., Wang, L.-F., Shi, Z.-L., and Zhou, P. (2018). Dampened STING-Dependent Interferon Activation in Bats. Cell Host & Microbe 23, 297-301.e4.

TABLES

(A)

Gene	Common name	Tajima's D
YGR098C	*ESP1*	-2.390412929
YLR397C	*AFG2*	-2.16261685
YHR023W	*MYO1*	-2.075765368
YMR168C	*CEP3*	-1.993925578
Enrichment *P* = 0.0263*		

(B)

Gene	Common name	Tajima's D
YGR098C	*ESP1*	-2.390412929
YKR054C	*DYN1*	-2.203644652
YLR397C	*AFG2*	-2.16261685
YDR180W	*SCC2*	-2.15471199
YHR023W	*MYO1*	-2.075765368
YMR168C	*CEP3*	-1.993925578
YNL172W	*APC1*	-1.841207567
YCR042C	*TAF2*	-1.791614606
Enrichment *P* = 0.0046*		

(C)

Gene	Common name	Tajima's D
YGR098C	*ESP1*	-2.390412929
YKL134C	*OCT1*	-2.293734473
YMR016C	*SOK2*	-2.281861163
YKR054C	*DYN1*	-2.203644652
YLR397C	*AFG2*	-2.16261685
YDR180W	*SCC2*	-2.15471199
YHR023W	*MYO1*	-2.075765368
YDR443C	*SSN2*	-2.070565734
YMR168C	*CEP3*	-1.993925578
YPR164W	*MMS1*	-1.990253695
YPL174C	*NIP100*	-1.967856781
YCR042C	*TAF2*	-1.791614606
YJR135C	*MCM22*	-1.651398253
YJL025W	*RRN7*	-1.502614461

Enrichment *P* = 0.0010*

(D)

Gene	Common name	Tajima's D
YBR136W	*MEC1*	-2.425141812
YMR275C	*BUL1*	-2.421881839
YDR103W	*STE5*	-2.336064982
YKL017C	*HCS1*	-2.301392848
YGR140W	*CBF2*	-2.268956348
YJR127C	*RSF2*	-2.193879044
YNL132W	*KRE33*	-2.185460334
YPL268W	*PLC1*	-2.179277149
YLR397C	*AFG2*	-2.16261685
YDR180W	*SCC2*	-2.15471199
YKL197C	*PEX1*	-2.146336148
YAL026C	*DRS2*	-2.131356095
YGL082W	*MIY1*	-2.118951228
YER151C	*UBP3*	-2.07655222
YDL035C	*GPR1*	-2.058601518
YDR508C	*GNP1*	-2.023301799
YDR235W	*PRP42*	-2.016907874
YLR422W	*DCK1*	-2.009696117
YJL062W	*LAS21*	-2.009120073
YIL068C	*SEC6*	-1.946442537
YDR456W	*NHX1*	-1.937222516
YGL095C	*VPS45*	-1.936591083
YNR045W	*PET494*	-1.908223035
YMR125W	*STO1*	-1.833598797
YCR042C	*TAF2*	-1.791614606
YNL049C	*SFB2*	-1.786844048
YOR092W	*ECM3*	-1.781980316
YML099C	*ARG81*	-1.772396018
YOR326W	*MYO2*	-1.746603341
YPL254W	*HFI1*	-1.722539604
YJR107W	*LIH1*	-1.718242115
YKL114C	*APN1*	-1.709523846
YMR078C	*CTF18*	-1.699771771
YPR049C	*ATG11*	-1.696994126
YMR167W	*MLH1*	-1.655296333
YOL081W	*IRA2*	-1.568684518

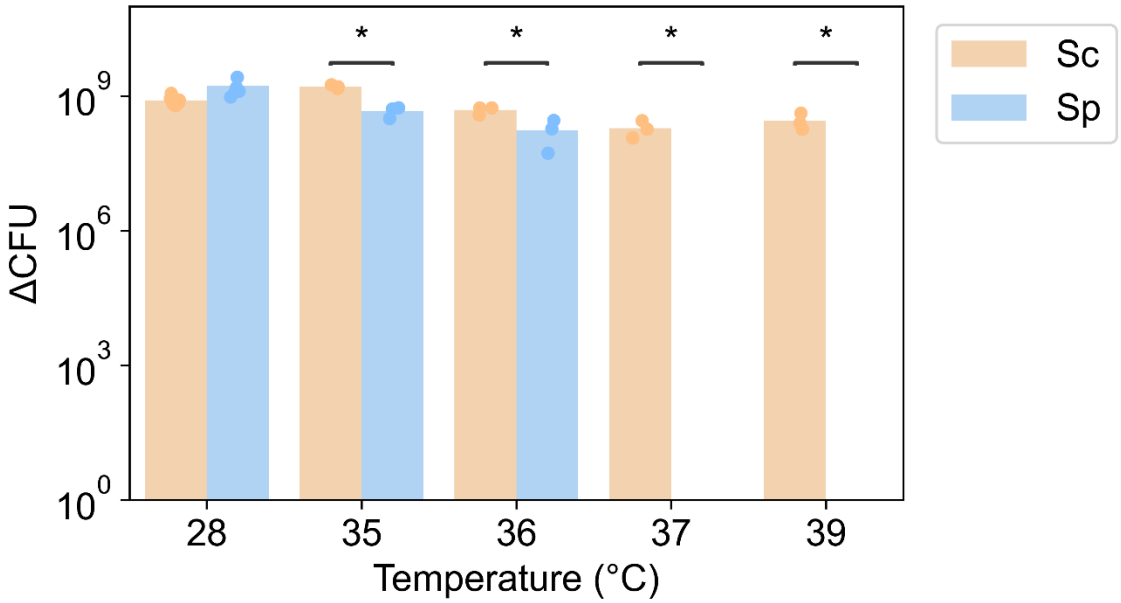
YOR091W	*TMA46*	-1.543998807
YMR207C	*HFA1*	-1.220103519
YIL152W	*VPR1*	-1.169622493
YMR094W	*CTF13*	-1.159736527
YDR375C	*BCS1*	-1.077792241
YDR318W	*MCM21*	-0.780309448
YOR371C	*GPB1*	-0.543345312
YGR198W	*YPP1*	-0.451850166
Enrichment *P* = 0.0012*		

Supplementary Table 1. Genes from thermotolerance mapping are enriched for low Tajima's D in *S. cerevisiae*. Each panel reports the analysis, in one set of hit genes from genetic dissection of thermotolerance between *S. cerevisiae* and *S. paradoxus,* of Tajima's D in genomes of wine/European strains of *S. cerevisiae* from (Peter et al., 2018). For a given panel, each of the first through penultimate rows reports the gene, common name, and Tajima's D. The final row reports the significance of the enrichment of low Tajima's D of the indicated genes in a genomic resampling test. (A) The four focal thermotolerance loci studied in this work. (B) The eight thermotolerance genes identified in (Weiss et al., 2018a). (C) The 14 thermotolerance genes identified in (Abrams et al., 2021a). (D) The 44 thermotolerance genes identified in (Abrams et al., 2021b). *, $P < 0.05$.

Species background	Strain background	Swap (if applicable)	Source	Strain Name
S. paradoxus	Z1	N/A	Weiss et al., 2018	CW62
S. cerevisiae	DBVPG1373	N/A	Weiss et al., 2018	CW68
S. cerevisiae	DBVPG1373	*S. paradoxus* Z1 CEP3 full swap	Weiss et al., 2018	CW73
S. cerevisiae	DBVPG1373	*S. paradoxus* Z1 AFG2 full swap	Weiss et al., 2018	CW64
S. cerevisiae	DBVPG1373	*S. paradoxus* Z1 MYO1 full swap	Weiss et al., 2018	CW104
S. cerevisiae	DBVPG1373	*S. paradoxus* Z1 ESP1 full swap	Weiss et al., 2018	CW98
S. cerevisiae	DBVPG1373	*S. paradoxus* N17 ESP1 full swap	Weiss et al., 2018	CW284
S. cerevisiae	DBVPG1373	*S. paradoxus* IFO1804 ESP1 full swap	Weiss et al., 2018	CW287

Supplementary Table 2. Strains used in this study.

FIGURES



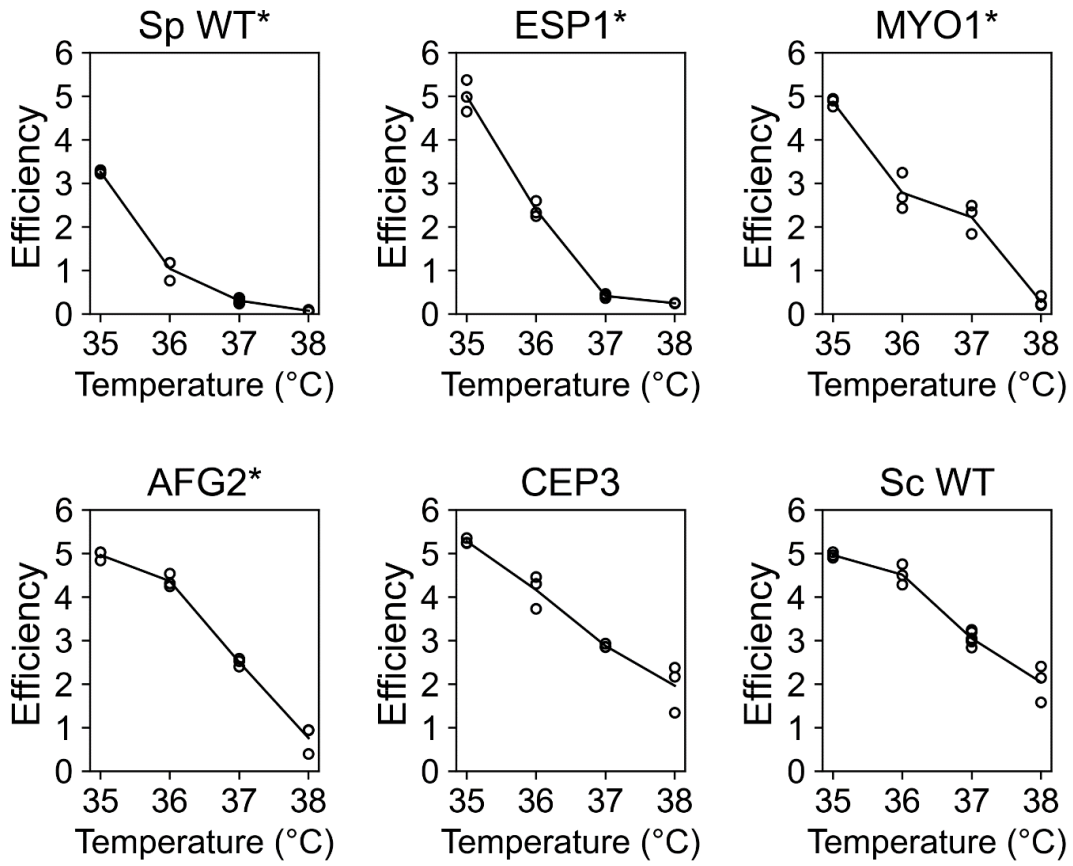


Figure 2. *S. paradoxus* alleles of thermotolerance genes confer temperature-dependent defects in the *S. cerevisiae* background. Each panel reports the results of growth experiments of a strain of *S. cerevisiae* DBVPG1373 harboring the indicated gene from *S. paradoxus* Z1, or the respective wild-type parent strains (WT), across a temperature gradient. In a given panel, the *y*-axis reports growth efficiency, the optical density reached by the culture after 24 hours at the temperature indicated on the *x*-axis, as a difference from the analogous quantity at time zero. Each point reports results from one biological replicate ($n = 3$), and the line represents the average growth efficiency of the indicated strain across the temperature gradient. *, $P < 10^{-3}$ for the strain by temperature interaction term of a two-factor ANOVA, in a comparison between the indicated strain and wild-type *S. cerevisiae*.

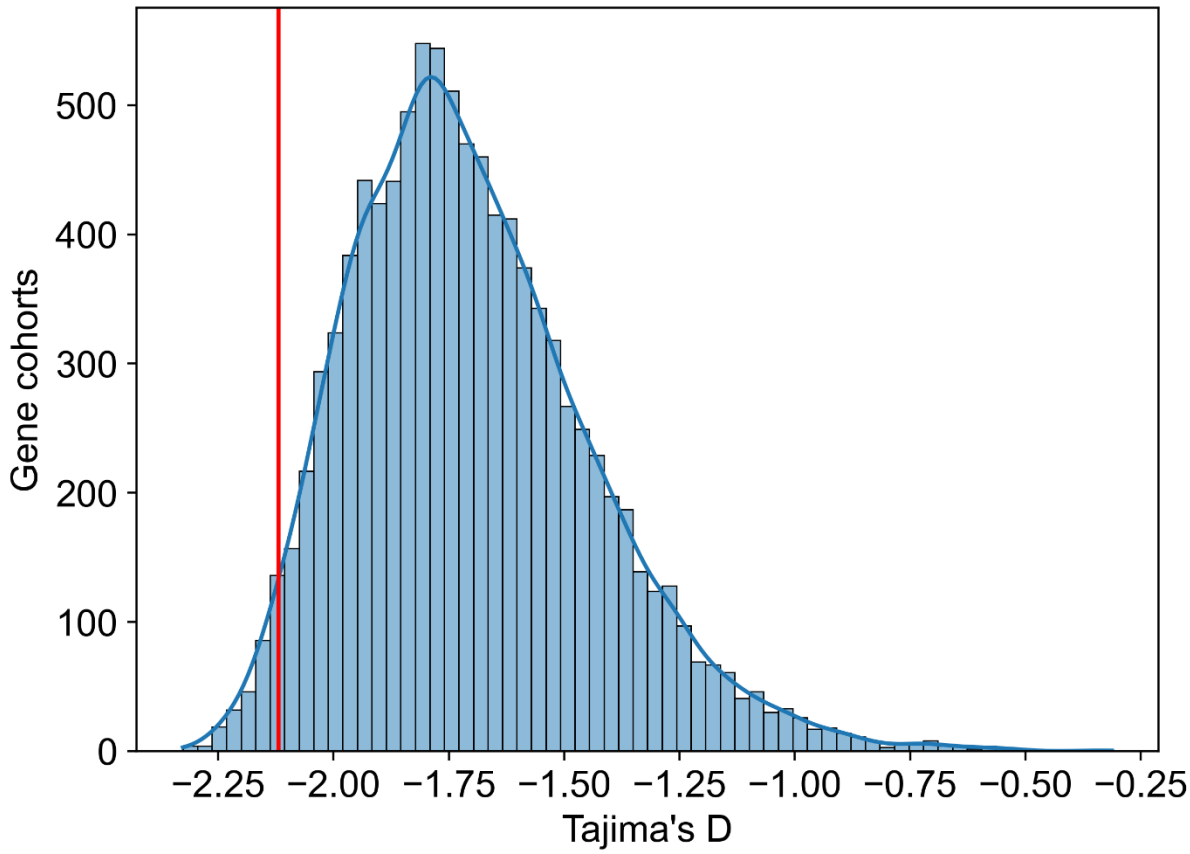
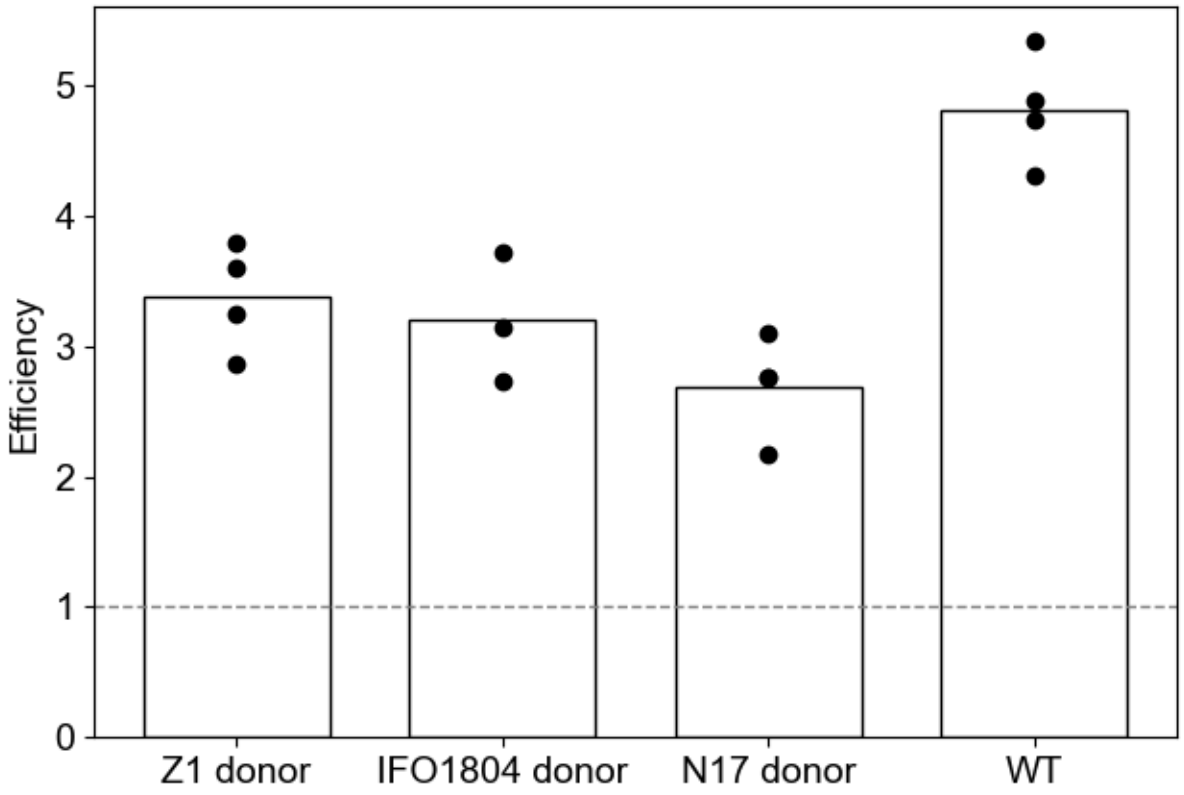


Figure 3. Thermotolerance genes are enriched for low Tajima's D in *S. cerevisiae*. The *x*-axis reports the median Tajima's D across a gene set of interest in genomes of wine/European strains of *S. cerevisiae* from (Peter et al., 2018). Blue bars reflect the results from genomic resampling, with the *y*-axis reporting the number of randomly chosen gene sets with the median Tajima's D shown on the *x*, and the blue curve showing a kernel density estimate of the histogram bar values. The red vertical line reports the Tajima's D value of the four thermotolerance genes characterized in Figure 2, corresponding to resampling $P = 0.0263$*. *, $P < 0.05$


Efficiency
0
1
2
3
4
5
Z1 donor
IFO1804 donor
N17 donor
WT

Chapter 5

Directed evolution leads to thermotolerant *S. paradoxus* with chromosomal duplications

ABSTRACT

Evolution builds traits over long periods of time, often employing complex, multigenic solutions to meet challenges to organismal fitness. Extant organisms can provide clues to the processes which may have occurred far back in evolutionary history—and also to aspects of the evolutionary landscape which nature either did not explore or did not favor during adaptation in the wild. Thermotolerance represents a putative adaptation in *Saccharomyces cerevisiae*, and its nearest relative, the thermosensitive *Saccharomyces paradoxus,* is a useful proxy for the ancestor of *S. cerevisiae.* In this chapter, we use *S. paradoxus* to study the evolutionary mechanisms by which thermotolerance can be gained under selection. Using a turbidostat as a morbidostat with respect to temperature, we evolved multiple lines of *S. paradoxus* which had a heritable thermotolerance phenotype. These evolved strains could grow at temperatures which are lethal to wild-type *S. paradoxus*, and next-generation sequencing revealed many independent instances of chromosomal duplication. These data shed light on an easily accessible evolutionary mechanism by which a thermosensitive *Saccharomyces* can rapidly acquire thermotolerance in a gradually increasing temperature paradigm.

INTRODUCTION

Saccharomyces cerevisiae is uniquely thermotolerant in its clade (Gonçalves et al., 2011; Hittinger, 2013; Salvadó et al., 2011; Sweeney et al., 2004; Weiss et al., 2018a) and its genome bears signatures of positive selection for this unique thermotolerance trait (Abrams et al., 2021a). Previous work in our lab and others has sought to uncover genetic determinants of this putative adaptation (Abrams et al., 2021b, 2021a; AlZaben et al., 2021; Weiss et al., 2018a). These works paint a picture of a multigenic and potentially omnigenic trait. Comparisons with *S. cerevisiae's* sister species *S. paradoxus* suggests that adaptation would have involved making modifications to many essential genes, including to essential gene coding sequences as in the case of the separase *ESP1* (Abrams et al., 2021b, 2021a; AlZaben et al., 2021; Weiss et al., 2018a). Our working model is thus that *S cerevisiae* attained a large number of variants that led to increased thermotolerance while preserving the indispensable function of many key housekeeping genes among those loci. We wanted to know how constrained the evolutionary landscape was to solve this complex problem.

Although modern *Saccharomyces* species serve as an example of a successful adaptation towards thermotolerance, potentially acquired over millions of years, they do not reveal what the initial landscape was for the acquisition of thermotolerance. Work in this thesis and elsewhere has characterized the independent, mostly beneficial allelic variation at unlinked loci that could have arisen along the *S. cerevisiae* lineage to fine-tune the thermotolerance trait over millions of years. In this chapter, we explore the idea that other molecular mechanisms could have been relevant at the earliest stages in

thermotolerance adaptation. Concepts such as the gene accordion provide a theoretical framework in which the initial evolutionary landscape lends itself to transient, and suboptimal adaptations that may not be preserved after further beneficial mutations arise (Elliott et al., 2013; Yona et al., 2012).

Landmark work in experimental evolution has revealed the critical roles of historical contingency (Blount et al., 2012), the order of acquisition of new mutations (Toprak et al., 2011), and chromosomal copy number (Blount et al., 2012; Yona et al., 2012) in early-stage adaptation. The relevance of these principles to thermotolerance acquisition in the wild in the *S. cerevisiae* lineage has remained unknown. Without access to the ancestor of modern-day *S. cerevisiae*, we decided to use the sister species *S. paradoxus* as a proxy for the ancestor and study it in its capacity as a thermosensitive *Saccharomyces* yeast. Directed evolution offered a powerful tool to look at how adaptation might play out in a population of thermosensitive *Saccharomyces* yeast faced with a selective pressure to acquire thermotolerance. We acknowledge that there is no way to fully know or replicate the conditions under which thermotolerance evolved in *Saccharomyces* in the wild. Instead, any lab-evolved strains offer insights into the kinds of heritable changes that would be available to make a yeast more thermotolerant. Our evolution experiments generated multiple independent strains of *S. paradoxus* with heritable increases in thermotolerance. We found evidence of aneuploidy in the form of chromosome duplications in our lab adapted strains, suggesting a possible mechanism by which this trait could be rapidly acquired in an environment with a strong selective pressure and abundant resources.

MATERIALS AND METHODS

Directed evolution

All directed evolutions were performed in a beta version of the eVOLVER turbidostat, developed by (Wong et al., 2018) and generously shared with us by the Arkin Lab. The turbidostat is now commercially available through Fynch Bio, with public code repositories on https://github.com/FYNCH-BIO. We modified the software from the beta version available on June 18, 2018, to create custom scripts which dynamically regulated the temperatures of the individual cultures evolving in parallel vials.

We modified the growth program so that the turbidostat would run as a morbidostat (Toprak et al., 2013). Briefly, the morbidostat program set the growth rate between a minimum and maximum threshold; if a population grew above 0.5 doublings per hour, the temperature of that vial increases by 0.1°C to apply selective pressure, but if the population growth rate dropped than 0.3 doublings per hour the temperature dropped by 0.1°C to prevent extinction.

We also modified the scripts to create separate, simpler temperature scenarios for our evolving cultures. We tested the response of *S. paradoxus* to static temperatures, without modifying the core code. Additionally, we created a version of the scripts which

ran a protocol to simply increase the temperature as a function of time (+0.1°C every 12 hours), instead of as a function of growth rate as above.

We inoculated all cultures for directed evolution with a wild-type *S. paradoxus* Z1 from the SGRP2 collection. A freezer stock was struck out on a YPD agar plate, and a single colony from that plate up was used to up inoculate a liquid pre-culture in YPD which was grown at 28°C and 200 RPM overnight before inoculating the 25-mL eVOLVER turbidostat continuous cultures. We found that eVOLVER vials with Pen-Strep were less likely to become contaminated over the course of the experiment, and liquid growth experiments showed no growth impact of the penicillin and streptomycin antibiotics on *S. paradoxus* growth as tested in in YPD at 28°C, 200 RPM, so we used YPD + Pen-Strep as the media following inoculation into the eVOLVER vials. Samples were collected from each evolution culture approximately weekly, at times of media change events, and stored at -80°C.

Modified eVOLVER beta scripts used in these experiments are available on GitHub at https://github.com/melanieabrams/bremdata

Next-Generation Sequencing (NGS)

We isolated gDNA from evolved cultures, using the Zymo YeaStar Genomic DNA Kit. Illumina library preparation and sequencing was performed by QB3 Genomics Sequencing Laboratory at UC Berkeley, to receive 10 Gb DNAseq data per sample from a pooled 150PE NovaSeq S4 flowcell.

NGS data analysis of evolved strains

NGS data analysis was performed to call variants and assess coverage change in our evolved strains. First, we subsampled our data with Seqtk (https://github.com/lh3/seqtk) by a factor of 100x or 500x from the larger, 10 Gb sample, to streamline the computation, using a pseudorandom seed of 100 so that mate pairs would align. Next, we used BBDuk (https://sourceforge.net/projects/bbmap) to trim Illumina adaptors and remove trace contaminants. Next, we trimmed reads for quality with SICKLE (https://github.com/najoshi/sickle/) using the default parameters for Illumina reads with Sanger-encoded phred33, and aligned reads which passed these filters to the reference genome as follows. We used RepeatMasker with RMBlast (http://www.repeatmasker.org) to mask repeat rich regions from the reference genome which would lead to large numbers of spurious alignments. Then we used Bowtie2 (Langmead and Salzberg, 2012) to index the masked reference genome and then map reads which passed quality control to that masked reference genome. With these mapped alignments in hand, we used Picard Tools (http://broadinstitute.github.io/picard/) to sort the SAM formatted alignments and convert them to BAM format. Then, we removed duplicate reads, and proceeded to call variants with freebayes and followed the initial variant calling with the freebayes VCF quality filter script with a cutoff of 20 (Garrison and Marth, 2012). We used custom scripts to filter the called variants based on depth and to search for genic variants and variants that were

neither present nor adjacent to variants also called in the wild-type sample. For 500x subsamples, we used a minimum read cutoff of 30 and maximum read depth cutoff of 200; for 100x subsamples, we used a minimum read cutoff of 300 and maximum read depth cutoff of 1250. We also sorted for a minimum allele balance of 0.3 to examine heterozygous variants. We then indexed sorted BAM files with SAMtools (Li et al., 2009) and used the Integrative Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) to visualize the variants and their context.

We assessed coverage changes in our NGS data as follows. We counted the coverage changes from the BAM files using the igvtools package of IGV. We then visualized the resulting TDF count files, and manually scanned for coverage changes using IGV.

Phenotyping evolved strains' thermotolerance

The thermotolerance of evolved strains was determined by a viability assay, as described in (Abrams and Brem, 2022). For the survey of viability phenotypes at high temperatures across wild-type isolates in Figure 2, strains were streaked out and a colony of each was pre-cultured in liquid as for 39°C growth above, except that the initial pre-culture to achieve saturation lasted 48 hours. Each pre-culture was back-diluted into 10 mL of YPD to reach an OD_{600}/mL of 0.05 and then cultured for 24 hours at the temperature of interest (28°C - 39°C). The viability of both the precultures and the cultures after 24 hours at the temperature of interest were measured with a spotting assay, where we diluted aliquots from the culture in a 1:10 series and spotting 3 µL of each dilution for growth on a solid YPD plate. After incubation at 28°C for two days, we used the dilution corresponding to the densest spot that was not a lawn for to determine viability: we counted the number of colonies in each of the two technical replicate spots, formulated the number of colony forming units per mL of undiluted culture (CFU/mL). We determined the change in viable cells by subtracting the number of cells in the culture at the initial timepoint from that at the final timepoint, based on the CFU/mL count and the culture volume**.**

RESULTS

Directed evolution of *S. paradoxus* to acquire *S. cerevisiae*-like thermotolerance.

We set out to conduct multiple parallel evolution experiments to investigate the evolutionary landscape of early adaptations towards thermotolerance in *Saccharomyces.* For this purpose, we chose the eVOLVER turbidostat, on the basis of its capacity for many small-scale cultures in parallel and its open codebase that we could modify to enable rising temperature to limit growth (Wong et al., 2018).

In our system, we wanted heat to be a primary selective pressure driving changes in the evolving yeast. The eVOLVER turbidostat dilutes the evolving cultures when the optical density of the cells within exceeds a certain threshold to maintain turbidity in a fixed range and prevent nutrient limitation; during these dilutions, half the culture volume containing cells would be discarded and replaced with fresh media. As a result, cells

with a greater rate of viable cell divisions would have a large selective advantage over their neighbors, in the repeated cycles of expansion and dilution. Thus, at a challenging temperature, a mutant which had overcome that challenge and had a greater rate of viable cell divisions could take over.

For our first set of experiments, we cultured the Z1 strain of *S. paradoxus,* a wild, diploid isolate from an oak tree niche in the UK, in our turbidostat at a fixed challenging temperature of 36°C. The results revealed an apparent “rescue” event where the growth rate suddenly increased to a new steady state after an initially plummeting growth rate, at approximately 100 hours, followed by a relatively constant growth rate over hundreds of hours (Figure S1A). In contrast, at a fixed, non-challenging temperature of 30°C, the growth rate did not plummet near the start of the experiment, and no such sudden increase in growth rate occurred at that time (Figure S1B). We conclude that our cultures of *S. paradoxus* cells could easily access mutational events at 36°C that allow them to improve fitness.

Because wild-type *S. cerevisiae* can grow at temperatures higher than 38°C, we next aimed to evolve *S. paradoxus* strains with a thermotolerance profile in that range. For this purpose, we first implemented a dynamic temperature program in which the temperature increased as a function of time, and cultured diploid Z1 *S. paradoxus* in the turbidostat under this regime. We again consistently saw an apparent “rescue” event early in the culture, at temperatures near 36°C, at which the growth rate suddenly increased after initially plummeting, at approximately 100 hours (Figure S2). However, after this point, as the temperature continued to steadily increase, the growth rate steadily declined, and ultimately the cultures failed to adapt further quickly enough for us to see any further similar “rescue” events that might suggest some change in the genetic or epigenetic state of the yeast within.

Given our goal of evolving *S. paradoxus* that tolerated high temperatures, we reasoned that a useful dynamic morbidostat control program toward this end would tie the temperature to the increasing apparent fitness of the yeast, with growth rate as the proxy for fitness. We implemented this with an automated approach that measured the growth-rate in real time, with the temperature increasing once the growth rate exceeded a threshold (see Methods), so that the evolution cultures could maintain a challenging temperature regime for even successful thermotolerant mutants. We inoculated diploid Z1 *S. paradoxus* cultures in a morbidostat running this setup and observed that, over time, in fits and starts, the temperatures where our evolved cultures could grow and divide well and quickly increased (Figures 1 and S2). The best-growing evolved strains of *S. paradoxus* grew at temperatures above 38°C, a temperature where wild-type *S. cerevisiae* but not wild-type *S. paradoxus* grow appreciably (Abrams and Brem, 2022; AlZaben et al., 2021; Weiss et al., 2018a).

Heritable changes in evolved strains of *S. paradoxus*

With adapted *S. paradoxus* in hand that could tolerate high temperatures, we next set out to gain insight into the molecular mechanisms at play. We first wanted to know

whether the growth-rate phenotypes which had guided our directed evolution would correspond to heritable changes in phenotype of the strains outside of the evolution cultures. We also wanted to know whether they would confer increased viability at higher temperatures, as we had theorized, rather than simply faster overall growth at a level which would compensate for the previously intolerable temperatures.

To test this, we carried out growth assays of adapted Z1 *S. paradoxus* isolates from frozen stocks sampled from our evolving cultures. The results revealed an increased viability in liquid cultures grown at 36°C, 200 RPM, a temperature where wild-type Z1 *S. paradoxus* loses more viable cells than it gains per unit of optical density over a 24-hour growth period in rich media (Figure 2). We conclude that thermotolerance as acquired by *S. paradoxus* in our experimental evolution was robust and heritable.

Genomic analysis of evolved strains of *S. paradoxus*

Next, we used next-generation sequencing to more deeply pursue the molecular mechanism underlying thermotolerance in the evolved strains of *S. paradoxus.*

We generated high-coverage genome sequencing data for four independently evolved strains of the Z1 *S. paradoxus* background. We called single nucleotide polymorphisms (SNPs) and found very few SNPs, with none shared across the four evolved strains. We narrowed our focus to non-telomeric loci, because short-read NGS contains a high proportion of errors in attempts to align repeat-rich regions, and telomeric regions are highly repeat-rich (Garrison and Marth, 2012; Hagemeijer et al., 2022; Mahajan et al., 2018). Only one SNP candidate, in the sequence encoding DNA damage checkpoint gene *DDC1*, corresponded to a codon change in a non-telomeric region, with a heterozygous mutation of Thr419Ile observed in one evolved strain (MA120). Sanger sequencing confirmed this mutation. To investigate the impact of this heterozygous variant on thermotolerance, we reasoned that the derived and ancestral alleles would segregate in a Mendelian manner during meiosis of the evolved strain, and potentially enable linkage analysis of the relationship between genotype and the thermotolerance phenotype. To pursue this, we dissected MA120, and isolated 15 viable progeny lines from four tetrads. We grew these progeny strains at 38°C, a temperature that the wild-type *S. paradoxus* cannot tolerate, and we observed complete viability in all cases. This result argued against a major, qualitative effect of the Thr419Ile mutation on thermotolerance in MA120, though we could not rule out a quantitative contribution by this variant.

Next, we investigated changes in copy number among our evolved thermotolerant strains of the Z1 *S. paradoxus* background. For this purpose, we analyzed read depth from our genome sequencing of each of four evolved strains as a proxy for copy number. The results revealed evidence for numerous duplications of various chromosomes in our evolved strains (Figure 3), many appearing in multiple replicates though none duplicated in all replicates. Since aneuploidies has been identified as a mechanism by which *S. cerevisiae* may increase its thermotolerance over even those levels seen in the thermotolerant wild-type in laboratory evolution (Yona et al., 2012),

we consider these duplications a strong candidate for the molecular mechanism by which our evolved *S. paradoxus* became thermotolerant.

DISCUSSION

When we study evolution in a laboratory setting, we gain insight into mechanisms by which evolution *could* build traits, in addition to those with which it *has* built such traits in the wild. Here, we show that the thermosensitive yeast *Saccharomyces paradoxus* can readily acquire thermotolerance on a timescale of months, at a level approaching the thermotolerance of *S. cerevisiae*. So, if the acquisition of thermotolerance in a *Saccharomyces* can be remarkably fast and straightforward, how come *S. cerevisiae* is the only member of its clade that can grow at such high temperatures? There are several possibilities.

First, it is possible that other *Saccharomyces* lineages did not face the same selective pressure as did the ancestor of *S. cerevisiae*. The latter is thought to have originated in a hot East Asian niche (Peter et al., 2018), and other ancestral populations that remained in a colder niche might not have been under pressure to acquire thermotolerance. In a potentially complementary model for the natural adaptation, it has been proposed that *S. cerevisiae* acquired thermotolerance and ethanol tolerance in tandem, since both heat and alcohol are a product of the Crabtree lifestyle and rapid growth which allow that yeast to dominate the niche of a fermenting fruit (Goddard, 2008; Salvadó et al., 2011).

Second, it is possible that pleiotropic effects of or trade-off mechanisms made the acquisition of thermotolerance unfavorable for wild *Saccharomyces*, even in the presence of selective pressure. There is evidence of such a trade-off between cryotolerance and thermotolerance at the thermotolerance loci in wild yeast (AlZaben et al., 2021), and other pleiotropic effects are possible too for both the molecular mechanisms that lead to thermotolerance in the wild and lab-evolved thermotolerant yeast.

Third, under a model in which chromosomal duplication was an initial event in the acquisition of thermotolerance in the wild, we can speculate that these duplications may have been too deleterious to be maintained in some lineages and/or environments. The lab-evolved strains of *S. paradoxus* from this work bear chromosomal duplications not seen in the wild yeast, and prior literature suggests that such duplications may have a heavy metabolic cost. In (Yona et al., 2012), transient aneuploidy reproducibly allowed lab-evolved *S. cerevisiae* to achieve increased thermotolerance, in that case through instances of trisomy in *S. cerevisiae* chromosome III. Those transient chromosomal duplications were later replaced with more metabolically efficient solutions. In our own work, our turbidostat kept ample nutrients available for evolving *S. paradoxus* by maintaining the cultures at a low density, so our strains never faced nutrient limitations as they adapted towards thermotolerance—even if the strains became less efficient. In the wild, a mechanism involving chromosomal duplications might not have been favorable if nutrients were scarcer. Modern *S. cerevisiae* does not have massive

chromosomal duplications; its thermotolerance adaptations involves alleles of many genes, including many essential genes (Abrams et al., 2021b, 2021a; Weiss et al., 2018a). This suggests that in any adaptation towards thermotolerance the ancestor of *S. cerevisiae* either had to acquire more refined solutions from the start, or that it made transient use of chromosomal duplications as it ultimately evolved thermotolerant alleles at many loci.

It is thus tempting to speculate that the results of an experimental evolution like ours would be different under nutrient limitation. We showed that operating a turbidostat as a morbidostat with respect to heat offers a fast and replicable method of creating heritably thermotolerant strains of the thermosensitive *S. paradoxus* yeast. A similar directed evolution experiment in a chemostat operated as a morbidostat could potentially reveal the molecular mechanisms of thermotolerance acquisition which are most readily accessible to a thermosensitive *Saccharomyces* in the presence of a selective pressure to remain metabolically efficient. Possibly, chromosomal duplications would also occur in this scenario, but would remain transient. If the ancestor of *S. cerevisiae* transiently acquired chromosomal duplications, the additional copies of genes on those chromosomes could permit protein evolution under relaxed selection leading to neofunctionalization with molecular functions that increased thermotolerance. Alternatively, more refined solutions like those developed by the ancestor of modern *S. cerevisiae* might become the most accessible way for yeast to overcome the challenge of thermotolerance in this constrained experimental setup. Although the ancestor of modern *S. cerevisiae* is long gone, experiments such as this, and those in this work, can offer insight into the selective forces and molecular mechanisms which shaped wild evolution. That, in turn, helps us understand the wild processes which shaped complex traits over long periods of time in the natural world around us.

DATA AVAILABILITY

Strains are available upon request. Custom scripts for operation of the eVOLVER turbidostat as a morbidostat with respect to temperature are available at https://github.com/melanieabrams-pub/eVOLVER_modified_codebase.

ACKNOWLEDGEMENTS

We thank the Arkin lab for their generosity with their eVOLVER turbidostat and computational resources, and in particular Harneet Rishi for his time and expertise with that equipment and its software. We thank Brandon Wong for his time and expertise with the eVOLVER turbidostat and codebase as well. We thank Patrick West for his time and expertise in variant calling in NGS data. We thank members of the Dueber and Brem laboratories for invaluable discussions.

REFERENCES

Abrams, M.B., and Brem, R.B. (2022). Temperature-dependent genetics of thermotolerance between yeast species.
Abrams, M.B., Dubin, C.A., AlZaben, F., Bravo, J., Joubert, P.M., Weiss, C.V., and Brem, R.B. (2021a). Population and comparative genetics of thermotolerance divergence between yeast species. G3 Genes|Genomes|Genetics.
Abrams, M.B., Chuong, J.N., AlZaben, F., Dubin, C.A., Skerker, J.M., and Brem, R.B. (2021b). Barcoded reciprocal hemizygosity analysis via sequencing illuminates the complex genetic basis of yeast thermotolerance.
AlZaben, F., Chuong, J.N., Abrams, M.B., and Brem, R.B. (2021). Joint effects of genes underlying a temperature specialization tradeoff in yeast. PLoS Genet 17, e1009793.
Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012). Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 489, 513–518.
Elliott, K.T., Cuff, L.E., and Neidle, E.L. (2013). Copy number change: evolving views on gene amplification. Future Microbiol 8, 887–899.
Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv:1207.3907 [q-Bio].
Goddard, M.R. (2008). Quantifying the complexities of Saccharomyces cerevisiae's ecosystem engineering via fermentation. Ecology 89, 2077–2082.
Gonçalves, P., Valério, E., Correia, C., Almeida, J.M.G.C.F. de, and Sampaio, J.P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric Saccharomyces Species. PLOS ONE 6, e20739.
Hagemeijer, Y.P., Guryev, V., and Horvatovich, P. (2022). Accurate Prediction of Protein Sequences for Proteogenomics Data Integration. In Clinical Proteomics: Methods and Protocols, F.J. Corrales, A. Paradela, and M. Marcilla, eds. (New York, NY: Springer US), pp. 233–260.
Hittinger, C.T. (2013). Saccharomyces diversity and evolution: a budding model genus. Trends Genet 29, 309–317.
Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
Mahajan, S., Wei, K.H.-C., Nalley, M.J., Gibilisco, L., and Bachtrog, D. (2018). De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. PLOS Biology 16, e2006348.
Peter, J., Chiara, M.D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556, 339–344.
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nat Biotechnol 29, 24–26.

Salvadó, Z., Arroyo-López, F.N., Guillamón, J.M., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus Saccharomyces. Appl Environ Microbiol 77, 2292–2302.

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

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14, 178–192.

Toprak, E., Veres, A., Michel, J.-B., Chait, R., Hartl, D.L., and Kishony, R. (2011). Evolutionary paths to antibiotic resistance under dynamically sustained drug stress. Nat Genet 44, 101–105.

Toprak, E., Veres, A., Yildiz, S., Pedraza, J.M., Chait, R., Paulsson, J., and Kishony, R. (2013). Building a Morbidostat: An automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. Nat Protoc 8, 555–567.

Weiss, C.V., Roop, J.I., Hackley, R.K., Chuong, J.N., Grigoriev, I.V., Arkin, A.P., Skerker, J.M., and Brem, R.B. (2018). Genetic dissection of interspecific differences in yeast thermotolerance. Nature Genetics 50, 1501.

Wong, B.G., Mancuso, C.P., Kiriakov, S., Bashor, C.J., and Khalil, A.S. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nature Biotechnology 36, 614–623.

Yona, A.H., Manor, Y.S., Herbst, R.H., Romano, G.H., Mitchell, A., Kupiec, M., Pilpel, Y., and Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. PNAS 109, 21010–21015.

FIGURES

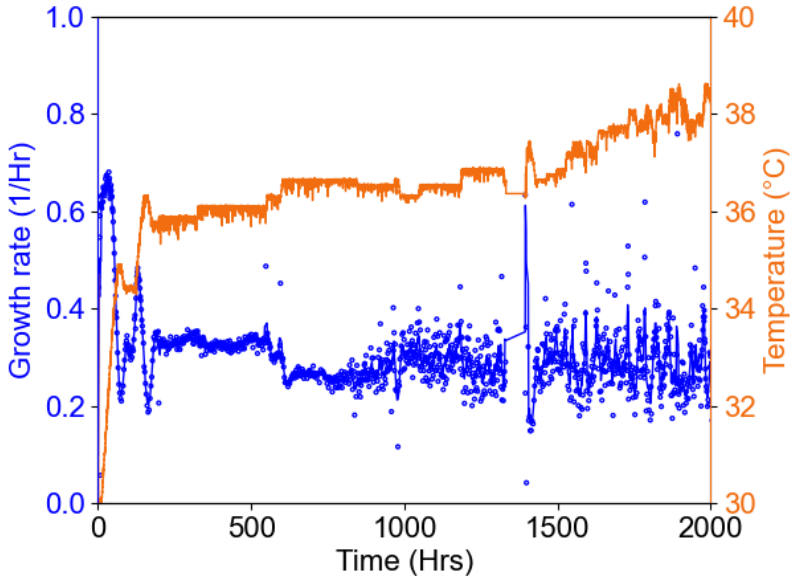


Figure 1. Directed evolution of *S. paradoxus* with a dynamically increasing temperature. The left *y*-axis and the light blue line represent the growth rate of a culture of *S. paradoxus* in the eVOLVER turbidostat; the right *y*-axis and a orange line represents the temperature, here increasing as a function of growth rate. The *x*-axis shows the time in hours for this experiment. A single vial's growth data is visualized for clarity; the "rescue" event at approximately 100 hours appeared in all tested replicates. Additional replicates are shown overlaid on the same plot in Figure S3.

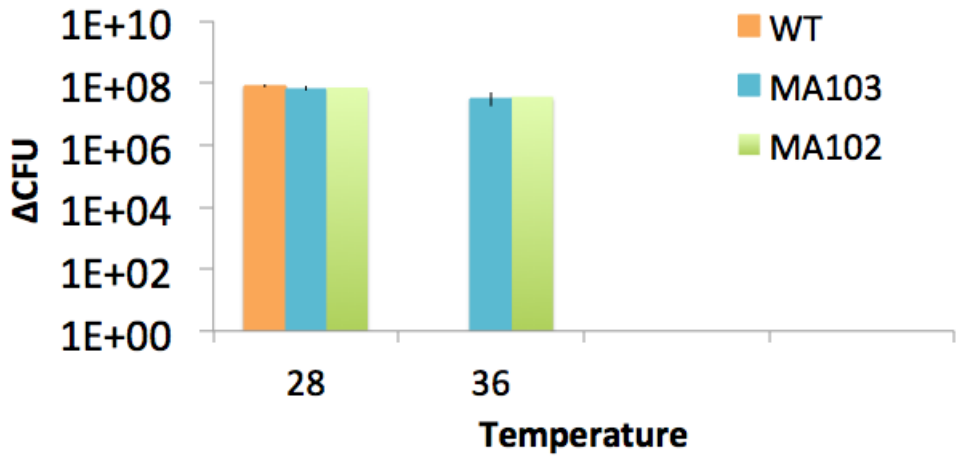


Figure 2. Heritable thermotolerance of evolved *S. paradoxus*. The *e*-axis shows strain viability, as assessed by change in the number of CFU per unit of OD of an *S. paradoxus* strain of the indicated genotype cultured over a 24-hour period; bars on the *e*-axis show the growth of the indicated strain at the indicated temperature. WT, wild-type; MA102 and MA103, strains evolved under dynamically increasing temperature.

A.

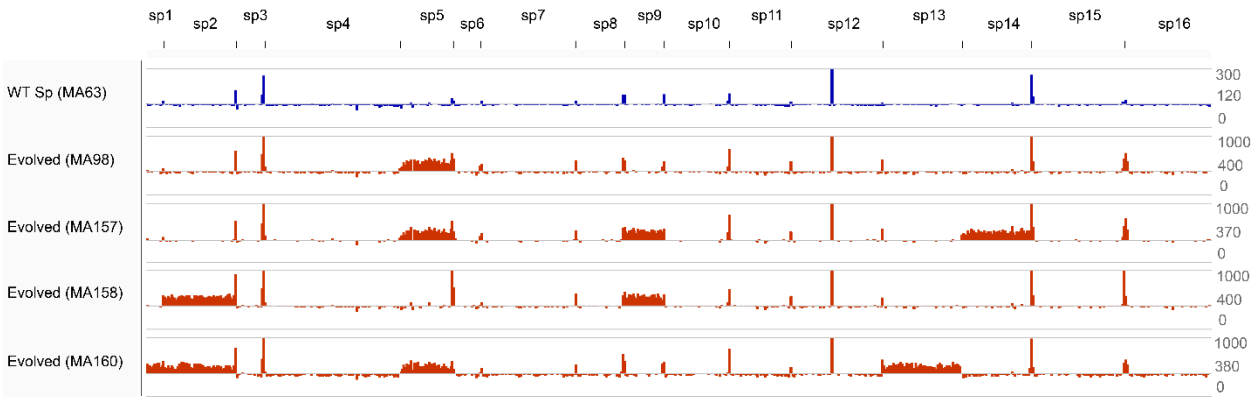
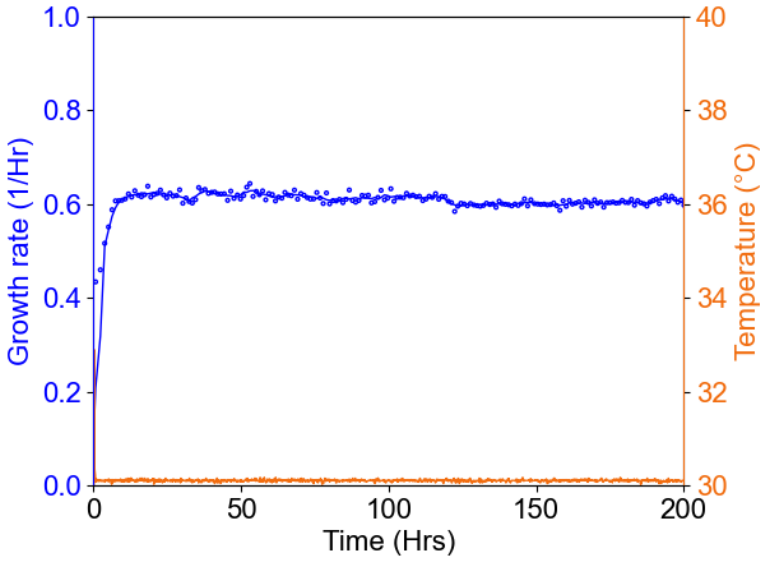


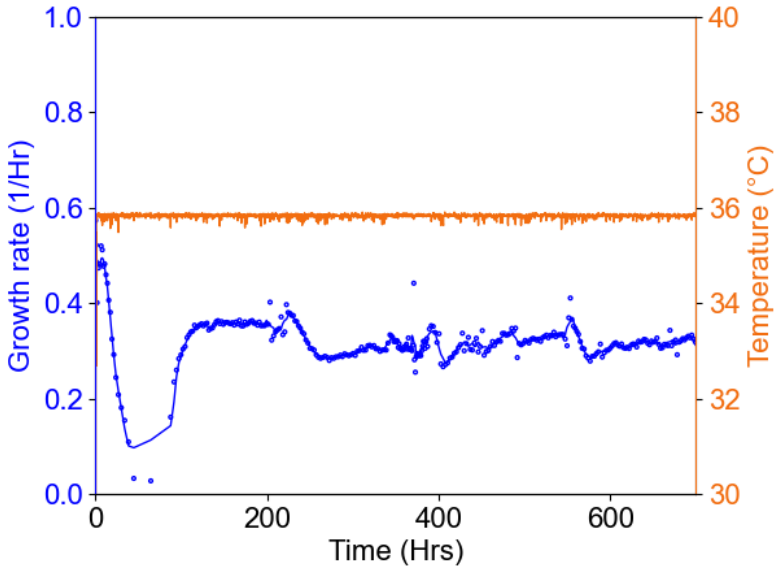
Figure 3. Coverage changes in genomes of evolved *S. paradoxus*. Each trace corresponds to the coverage of the *S. paradoxus* Z1 genome by the mapped wild-type *S. paradoxus* (blue) or the indicated strain evolved under dynamically increasing temperature (red). The maximum, minimum, and zero-point of the data range for each trace is indicated in the grey labels to the right of each trace.

SUPPLEMENTARY FIGURES

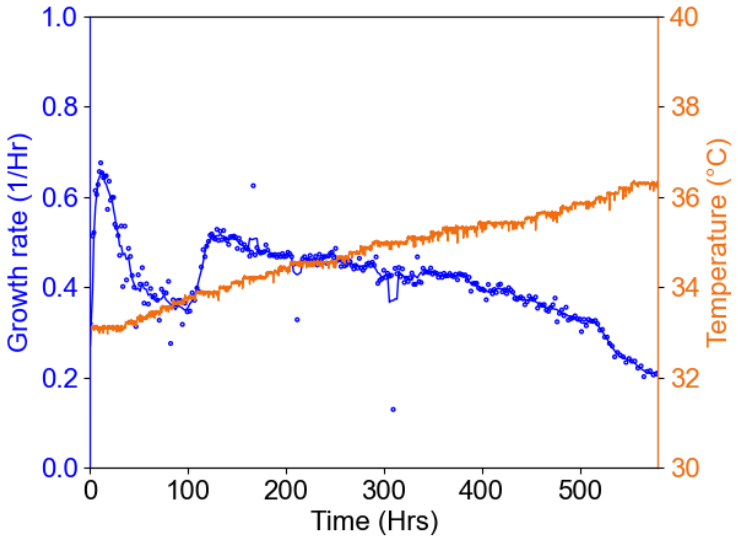
A.



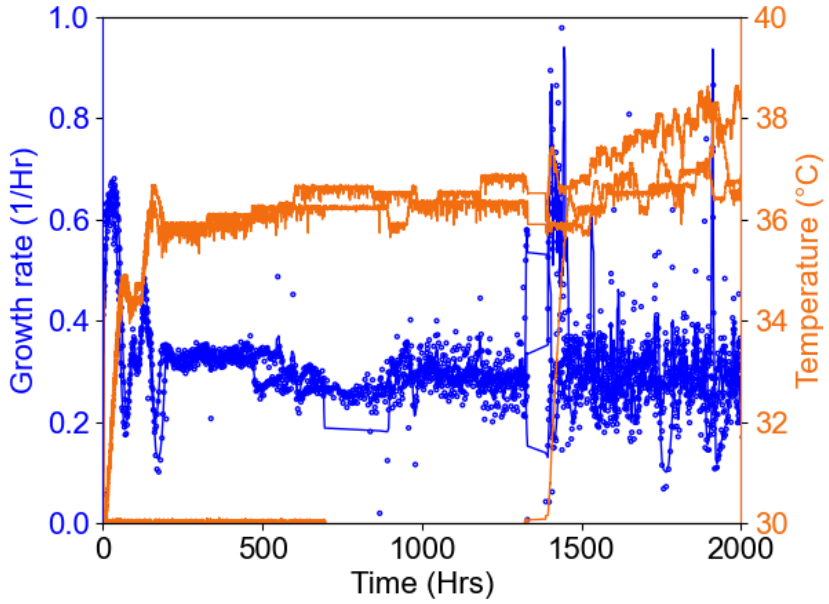
B.



Supplementary Figure 1. Directed evolution of *S. paradoxus* with a fixed temperature program. Each panel reports results from a culture of *S. paradoxus* in the eVOLVER turbidostat at a fixed temperature. In each, the left *y*-axis and the dark blue line represent the growth rate of a culture of *S. paradoxus* in the eVOLVER turbidostat; the right *y*-axis and the orange line represent the temperature. A. Fixed temperature at 30°C. Fixed temperature at 36°C.



Supplementary Figure 2. Directed evolution of *S. paradoxus* with temperature increasing as a function of time. The left *y*-axis and the light blue line represent the growth rate of a culture of *S. paradoxus* in the eVOLVER turbidostat; the right *y*-axis and a orange line represents the temperature, here increasing as a function of time. The *x*-axis shows the time in hours.



Supplementary Figure 3. Growth rate and increasing temperature of parallel eVOLVER cultures with dynamic temperature program. The data are as in Figure 1 of the main text except that the growth rates and temperatures for three separate culture vials are shown, with one line plotted for the growth rate and temperatures of each vial. The sustained spike in growth rate around 1400 hours in one culture corresponds to a contamination event after which that particular culture was restarted from a freezer stock of the last pre-contamination timepoint.