Molecular and evolutionary mechanisms of yeast thermotolerance

Melanie B. Abrams[1]*, Claire A. Dubin[1]*, Faisal AlZaben[1], Juan Bravo[2], Pierre M. Joubert[1], Carly V. Weiss[1,3], and Rachel B. Brem[1,4]

[1]Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA; [2]Graduate Program in the Biology of Aging, University of Southern California, Los Angeles, CA; [3]Current address: Department of Biology, Stanford University, Palo Alto, CA; [4]Buck Institute for Research on Aging, Novato, CA

*These authors contributed equally to this work.

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 growth advantage of the yeast *Saccharomyces cerevisiae* at high temperature, relative to its sister species *Saccharomyces paradoxus*, as a model for addressing these questions. Starting with a suite of identified loci where the *S. cerevisiae* allele promotes thermotolerance, we detected robust signatures of positive selection with sequence-based molecular evolution tests. These included recurrent signal across dozens of *S. cerevisiae* populations; we thus propose an approximate date for a selective sweep of thermotolerance alleles in the ancestor of the modern lineages of the species. In focused experiments on the chromosome segregation gene *ESP1*, a particularly strong contributor to thermotolerance, the *S. paradoxus* allele conferred a defect in cell division at high temperature, relative to the *S. cerevisiae allele*. Only divergent sites in the *ESP1* coding region mattered phenotypically, with no functional impact detectable from the promoter. This study of an adaptation that likely arose long ago, by sweeps at many unlinked loci, sets the groundwork for evolutionary genetics in ancient traits across Eukarya.

INTRODUCTION

A central goal of research in evolutionary genetics is to understand how new traits are built. Much of the literature to date has focused on putatively adaptive trait variation 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 the mechanisms of adaptation over relatively short timescales. Recent insights include the importance of the order in which adaptive alleles are acquired, when they interact epistatically (Blount et al., 2012; Toprak et al., 2011); the coexistence of multiple equally fit genotypes in a population (soft selective sweeps; (Barroso-Batista et al., 2014; Castro et al., 2019; Garud et al., 2015; Good et al., 2017; Harris et al., 2018)); and the origin of adaptive alleles as standing variants, pre-existing in populations before the selective event (Xie et al., 2019).

This body of work on the mechanisms of recent, or local, 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, and their genetics can be hard to track down. Candidate-gene studies have implicated individual loci in trait differences between species (Anderson et al., 2016; Baldwin et al., 2014; Liu et al., 2018; Massey and Wittkopp, 2016; Sackton et al., 2019; Sulak et al., 2016; Tian et al., 2019). And landmark cases have reconstructed the path taken by ancient evolution in repacking a particular protein of interest (Anderson et al., 2015; Bridgham et al., 2009; Finnigan et al., 2012; Liu et al., 2018; Pillai et al., 2020). However, most traits are likely to be genetically complex; and to date, as a field we have had little insight into the mechanisms by which evolution coordinates the rise of adaptive alleles all over the genome.

In prior work, our group mapped multiple housekeeping genes underlying an adaptive thermotolerance trait that distinguishes *Saccharomyces cerevisiae* as a species from its closest relative, *S. paradoxus* (Weiss et al., 2018). Here we set out to infer when and how adaptive alleles were acquired at these genes, by testing for signatures of natural selection in *S.*

cerevisiae populations. Then, we used the results as a jumping-off point for additional analyses of the molecular mechanisms by which variants at thermotolerance loci confer their effects.

RESULTS

Pervasive divergence from *S. paradoxus* at *S. cerevisiae* thermotolerance loci

We started by addressing the genetics of *S. cerevisiae* thermotolerance, with the ultimate goal of investigating evolutionary mechanisms underlying the trait. Our earlier study used genome-scale screens with the reciprocal hemizygosity test (Steinmetz et al., 2002; Stern, 2014) to identify genes at which *S. cerevisiae* harbored pro-thermotolerance alleles relative to those of *S. paradoxus* (Weiss et al., 2018). We re-processed these screen data with an improved statistical workflow, to boost statistical power and genome coverage (see Methods). The results yielded ten loci at which variation was associated with thermotolerance (Table S1), of which seven were reported and validated in our original report. We considered the set of ten 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 character of *S. cerevisiae*, shared among isolates within the species and distinguishing them from *S. paradoxus* and other members of their clade (Gonçalves et al., 2011; Salvadó et al., 2011; Sweeney et al., 2004). We hypothesized that sequence-based signatures of natural selection would be detectable across *S. cerevisiae* population at loci that underlie the thermotolerance trait. We used the absolute diversity statistic D_{xy} for an initial test of this notion. D_{xy} reaches high levels in a lineage after selection when compared to a representative of the ancestral state (Nei, 1987). We applied this metric to genomes from a large, recent survey of *S. cerevisiae* populations (Peter et al., 2018) and from the deepest-sampled *S. paradoxus* populations available (one from vineyards and European collection locales (Bergström et al., 2014) and another from North America (Durand et al., 2019)). We found enrichment for high D_{xy} among our mapped thermotolerance genes in each population pair in turn (Table S2). Thus, divergence from *S. paradoxus* at thermotolerance loci is a trend that pervades *S. cerevisiae*, attesting to the likely evolutionary and functional relevance of these genes across the species.

At thermotolerance genes, common alleles within *S. cerevisiae* are particularly common

Next, we investigated variation among isolates of *S. cerevisiae* at thermotolerance genes. If the trait arose under selection, an adaptive allele at any locus would manifest as a shared haplotype, *i.e.* a block of sequence homogeneity across members of a population (Berry et al., 1991; Smith and Haigh, 1974; Weigand and Leese, 2018). We used SelectionHapStats (Garud et al., 2015; Harris et al., 2018) to quantify frequencies of common multi-locus genotypes (MLGs; analogous to haplotypes, for unphased genotype data) in an *S. cerevisiae* population, across sequence windows. We anticipated that large genomic data sets would afford the best power for these analyses, and as such we focused on the five *S. cerevisiae* populations with the deepest sampling (Peter et al., 2018). In each population in turn, we first tabulated the most common MLG at a given locus. We found our thermotolerance genes to be enriched for high values of allele frequency in all tested *S. cerevisiae* populations, significantly so in three cases (Table 1 and Table S3A). Thus, at a thermotolerance locus the most common allele tended to be particularly common, and this trend was detectable across *S. cerevisiae* as a species. We also tabulated the frequency of the second most common MLG relative to the first, which when elevated serves as a hallmark of a soft selective sweep (Messer and Petrov, 2013). We detected no signal for high values of this ratio at thermotolerance loci. Rather, in several

populations, the second most common MLG was markedly low in frequency for the thermotolerance gene cohort (Table 1 and Table S3B). These findings point to hard selective sweeps as a better description of thermotolerance locus evolution than the soft-sweep model, as a general rule across the gene set and across populations.

We reasoned that inspecting MLG frequency at high spatial resolution could help reveal facets of the selective history of our loci. As a first resource to explore this, we made use of the most deeply sampled *S. cerevisiae* population, that of vineyard and European isolates (362 strains; (Peter et al., 2018)). In this set of genomes, we found that the frequency of the most common MLG was not uniform across a given thermotolerance gene, and for most loci, peaks of high allele frequency could be resolved (Figure 1A and Figure S1). As expected, such peaks corresponded to troughs of polymorphism across the population, measured by the number of pairwise differences between strains (π), and likewise low allele-sharing correlated with high polymorphism (Figure S1).

Notably, even at peaks of allele-sharing in thermotolerance loci in wine/European *S. cerevisiae*, in absolute terms allele frequencies were modest. In the top-scoring regions in the wine/European population, at the 5' end of the chromosome segregation gene *ESP1* (Figure 1A), at most 15% of strains harbored the most common MLG. These data do not support the inference of a recent hard selective sweep of a single adaptive allele, after which we would expect tight conservation for long genomic stretches. Most parsimoniously, partial allele-sharing at the 5' end of *ESP1* can be interpreted as the residual signal of an ancient hard sweep, eroded by mutation, recombination, and/or demographic effects. We observed similar spatial patterns at *ESP1* in the other *S. cerevisiae* populations in which we had detected allele-sharing in general across thermotolerance genes (Figure 1B-C), indicating that all these populations likely have had the same forces at play at the *ESP1* locus.

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

The maximal signal of allele-sharing we detected at *ESP1* within multiple sampled populations *S. cerevisiae* (Figure 1, Table S1A) dovetailed with its distinction, among thermotolerance genes, as the one with the largest effect on growth at high temperature in swaps between *S. cerevisiae* and *S. paradoxus* (Weiss et al., 2018). Given this evidence for evolutionary and functional importance of variation at *ESP1*, we chose to use it as a test case to achieve deeper molecular insight in the thermotolerance system. We first aimed to contrast the role of coding and *cis*-regulatory divergence between species at *ESP1*. We used the branch-site model of the PAML suite (Yang, 2007) for phylogenetic analyses of the *ESP1* coding sequence in *Saccharomyces sensu stricto* species. The results revealed significant evidence for accelerated protein evolution along the *S. cerevisiae* lineage of *ESP1* ($p = 0.046$), though we could not resolve signal at individual codons (see Methods). Thus, *S. cerevisiae* harbors unique amino acid changes at sites otherwise conserved across sister species.

This molecular-evolution signature in the *ESP1* open reading frame suggested that its protein divergence between species would have particularly strong impact on thermotolerance. We tested this experimentally by introducing the *ESP1* coding region and, separately, the *ESP1* promoter, from wild *S. paradoxus* (the UK isolate Z1) into a wild *S. cerevisiae* background (the Dutch strain DVBPG1373). Growth experiments revealed a dramatic, temperature-dependent effect of interspecies divergence at the *ESP1* open reading frame, with the *S. paradoxus* allele compromising growth at high temperature (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, species divergence

in the *ESP1* promoter had no effect on thermotolerance, when analyzed on its own (Figure 2 and Figure S2). As an independent analysis of potential promoter effects, we inspected *cis*-regulatory variation between *S. cerevisiae* and *S. paradoxus* in measurements of *ESP1* gene expression. We detected no striking *cis*-regulatory divergence, at *ESP1* in particular or across thermotolerance genes as a set (Table S4). Likewise, the latter yielded no signal in a test for non-neutral directional *cis*-regulatory change ((Bullard et al., 2010); Table S4). These 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 prove to 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., 2018), and their relevance to growth under other conditions is unknown. Using *ESP1* as a case study to address this question, we assayed biomass accumulation of wild-type and *ESP1* transgenic strains under a temperature dose-response. In these growth experiments, we observed a gradual decline in *S. cerevisiae* and *S. paradoxus* growth as temperature increased, with the latter more sensitive to heat as expected (Figure 3). An 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., 2018), 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 is compromised under heat treatment, blocking the progress of the cell cycle and limiting biomass accumulation.

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. In this work, we used yeast thermotolerance as a model of this process. We shed light on how, why, and when the trait was acquired in *S. cerevisiae*. We also established key facets of the molecular genetics and cell biology of divergent alleles at the underlying loci.

Our sequence analyses yielded a clear picture of the selective forces driving thermotolerance: we found robust signatures of positive selection at the causal loci, within and between species.

Most striking was the agreement in these results between *S. cerevisiae* populations from around the world, for thermotolerance loci as a group and when we focused on the spatial breakpoints of allele-sharing at *ESP1*. This pervasive molecular-evolution signal across *S. cerevisiae* argues against a history of recent repeated sweeps per locus in the different lineages of this species, of the kind characterized in elegant studies of local adaptation (Chan et al., 2010; Hoekstra and Nachman, 2003; Rosenblum et al., 2010; Xie et al., 2019). Rather, for yeast thermotolerance, the most parsimonious model is one of a single sweep at a given gene in the ancestor of modern-day *S. cerevisiae* populations, before the radiation of the latter. Plausibly, the initial rise of the trait could have been driven by the ecology of hot East Asian niches where *S. cerevisiae* likely originated (Peter et al., 2018). Our sequence analyses would then pick up only the remnants of signal from swept alleles maintained in extant genomes. Thus, for example, we would interpret the rise and fall in allele-sharing we detect across *ESP1* as a product of spatial differences in mutation and recombination, as recent variants acumulate on a haplotype that swept long ago.

That said, having inferred the model of ancient hard sweeps largely from features of thermotolerance loci taken on average across their cohort, we acknowledge that any one gene may have followed a different evolutionary path. Indeed, we are likely missing much of the genetics of *S. cerevisiae* thermotolerance in the first place. Our original mapping screen of the trait, which used European *S. cerevisiae,* was likely subject to false negatives. And of the genes we did identify in that case study, some may have no role in the thermotolerance genetics outside the European population. But our ability to detect molecular-evolution signal in the face of these caveats attests to the power of the gene-cohort approach.

Our data leave open the question of which divergent sites at a given gene underlie *S. cerevisiae* thermotolerance. A primary role for protein-coding variation, as we have seen for *ESP1*, would echo decades of evo-devo literature on this theme (Hoekstra and Coyne, 2007; Wray, 2007). Furthermore, 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). Future work will be necessary to test whether *S. cerevisiae* truly has tweaked its protein biophysics to meet the challenges of growth at high temperature, and how this might compare with the evolution of thermotolerance in other systems (Sas-Chen et al., 2020).

In fact, our work has set a precedent by which evolutionary-genetic principles for many other ancient traits may come within reach. The paradigm we have used here combines methods for inter-species genetics (Weiss and Brem, 2019) with molecular-evolution tools. This setup is generally applicable across Eukarya; and with it, once we know the genes that govern a species-defining trait, we can look far back in history and figure out how it came to be.

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., 2018) as described, with the following exceptions. Call $a_{39,i}$ the average, across technical replicates, of sequencing-based abundances of a hemizygote mutant measured after 5-6 generations in biological replicate *i* of growth at 39°C, and $a_{28,i}$ the analogous quantity for growth at 28°C. We tabulated three replicate estimates of the temperature effect on growth of the mutant as $\log2(a_{39,i}/a_{28,i})$ for i = (1,3). If the coefficient of variation across these biological replicates was greater than 20.0, 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_{Scer}. We did likewise for the *S. cerevisiae* allele, yielding s_{Spar}. We retained for further analysis only genes at which either of these two vectors was of length two or greater. For each such gene, we compared s_{Scer} and s_{Spar} with a Wilcoxon test, and corrected for multiple testing across genes, as described. Instructions and scripts are available from the authors.

Sequence data, alignments, and interspecies diversity

For D_{XY} analyses in Table S1, 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). 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 population.

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, we first tabulated D_{therm}, the median D_{xy} across the thermotolerance gene cohort. 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 D_{xy} in each D_{rand}. We used as an empirical *p*-value the proportion of random cohorts with $D_{rand} \geq D_{therm}$.

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

We calculated G1, the frequency of the most common multi-locus genotype, and G2/G1, the relative frequencies of the second and first most common multi-locus genotypes, as follows. We used unphased VCF genotypes for each of the five largest *S. cerevisiae* populations 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. For each gene in the genome for which we had genotype calls, for a given population we tabulated G2/G1 in each window whose center fell within the open reading frame, and we calculated the average across the windows.

In Table 1, to assess whether thermotolerance genes had a G2/G1 distinct from those in the genome as a whole, we first calculated $G2/G1_{therm}$, the median across our ten thermotolerance loci. 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 G2/G1, $G2/G1_{rand}$. We then used as an empirical *p*-value the proportion of random cohorts with $G2/G1_{rand} \geq G2/G1_{therm}$, multiplying by 2 to render the test two-tailed. We followed the same strategy to calculate and assess enrichment of G1 in thermotolerance genes, except that we used a one-tailed resampling test. Separately, to check for effects of gene length, we devised a distinct analysis strategy in which random cohorts were resampled with the same length distribution as that of the true thermotolerance gene cohort; results for G1 in the Wine/European *S. cerevisiae* population mirrored those from essentiality-based resampling (data not shown).

Polymorphism in Wine/European *S. cerevisiae*

For the Wine/European *S. cerevisiae* population from (Peter et al., 2018), we used genotype data as a VCF as input into VCFtools (Danecek et al., 2011) with the command --site-pi. This output a polymorphism (π) value for each SNP. For Figure S1, we tabulated the average π across a 1200 SNP window centered at each SNP, with invariant sites contributing $\pi = 0$.

Custom scripts for sequence preparation and nucleotide diversity calculations are available at https://github.com/clairedubin/thermotolerance.

***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. 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*

In Table S4, we used measurements of high- and standard-temperature allele-specific expression in an *S. cerevisiae* x *S. uvarum* hybrid from (Li and Fay, 2017), and in an *S. cerevisiae* x *S. paradoxus* hybrid from (Tirosh et al., 2009), as estimates of *cis*-regulatory variation between the respective species pairs at each thermotolerance gene.

Separately, we tested for directional, temperature-dependent *cis*-regulatory change across the thermotolerance gene cohort as follows. Using the data from (Li and Fay, 2017) for each gene in turn, given the allele-specific expression measurement $a_{T,Scer}$ from the *S. cerevisiae* allele at temperature *T* (37°C or 33°C), or $a_{T,Suv}$ calculated analogously for the *S. uvarum* allele, we tabulated the temperature-dependent allele-specific expression ratio ASE = $\log(a_{37,Scer}/a_{37,Suv}) - \log(a_{33,Scer}/a_{33,Suv})$. We took the median of this quantity across all thermotolerance genes 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 ASE, ASE_{rand}. We then used as an empirical *p*-value the proportion of random cohorts with $ASE_{rand} \geq ASE_{therm}$. Separately, we did the analogous calculation with the measurements from cultures at 35°C and 30°C from (Tirosh et al., 2009).

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., 2018). Strains are listed in Table S5.

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., 2018), 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 strain 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. We used the complete cohort of measurements of $e_s/\langle e_{Scer} \rangle$ from all days 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 4, 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., 2018) 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., 2018). 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.

ACKNOWLEDGEMENTS

The authors thank J. Chuong for verifying and maintaining transgenic strains, D. Savage for his generosity with lab facilities and resources, and J. Schraiber for discussions. This work was supported by NSF GRFP DGE 1752814 to M.A. and NIH R01 GM120430 to R.B.B.

BIBLIOGRAPHY

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.

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 Genet. *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.

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 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.

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.

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.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

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.

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.

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.

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.

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*.

Messer, P.W., and Petrov, D.A. (2013). Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol. Evol. (Amst.) *28*, 659–669.

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.

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.

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.

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.

experiment for longer limbs in mice. Elife *8*.

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 *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. Proc. Natl. Acad. Sci. U.S.A. *111*, 13912–13917.

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.

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 Genetics *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 Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evolution *61*, 995–1016.

Hoekstra, H.E., and Nachman, M.W. (2003). Different genes underlie adaptive melanism in different populations of rock pocket mice. Mol. Ecol. *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*.

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.H., and Wittkopp, P.J. (2016). The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species. Curr. Top. Dev. Biol. *119*, 27–61.

Nei, M. (1987). Molecular Evolutionary Genetics (Columbia University Press).

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.

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. Proc. Natl. Acad. Sci. U.S.A. *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 *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. Genetics 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.

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.

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.

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., and Brem, R.B. (2019). Dissecting Trait Variation across Species Barriers. Trends Ecol. Evol. (Amst.) *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.

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.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

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. Mol. Biol. Evol. *24*, 1586–1591.

TABLES AND FIGURES

		G1			G2/G1		
Population	**Number of Isolates**	**Thermotolerance genes**	**Genome**	**p-value**	**Thermotolerance genes**	**Genome**	**p-value**
Wine/European	362	0.048	0.026	0.0101*	0.362	0.602	0.0484*
Mosaic Region 3	113	0.022	0.019	0.0492*	0.803	0.819	0.669
Mixed Origin	72	0.077	0.061	0.0677	0.688	0.715	0.290
Sake	47	0.278	0.257	0.0608	0.158	0.130	0.609
Brazilian Bioethanol	35	0.133	0.109	0.0284*	0.263	0.363	0.0456*

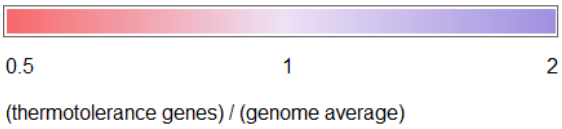
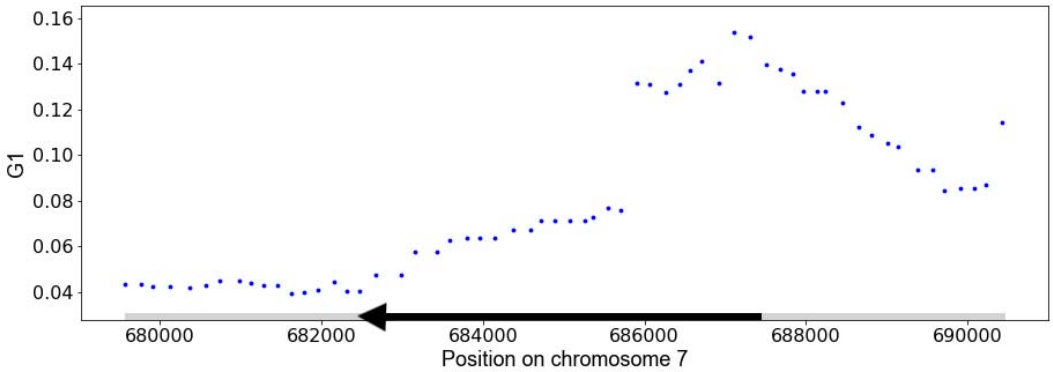
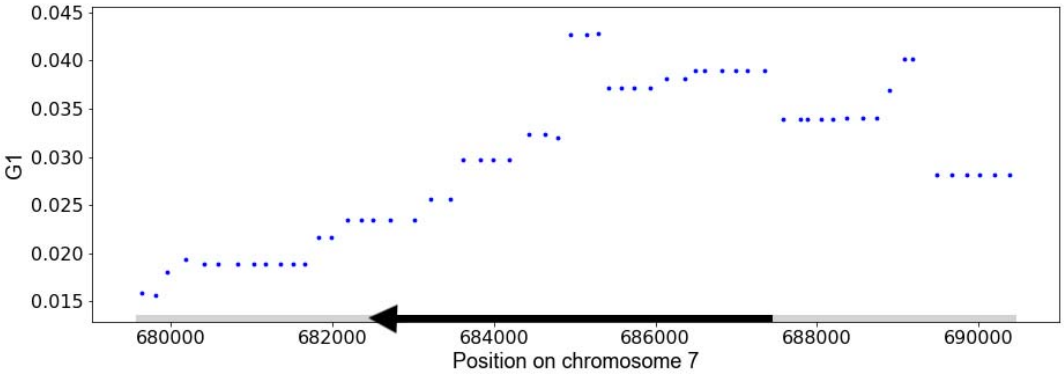


Table 1. Thermotolerance loci are enriched for allele-sharing within *S. cerevisiae* populations. Each row reports results from analyses of multi-locus genotypes (analogous to haplotypes for unphased data; (Harris et al., 2018)) at thermotolerance loci in the indicated population of *S. cerevisiae* from (Peter et al., 2018). In a given population, the third and fourth columns report the median frequency of the most abundant genotype, G1, in thermotolerance genes and random sets of genes from the genome, respectively; the fifth column reports empirical significance from a resampling test. The sixth through eight columns report analogous results for the ratio of the second-most abundant genotype to the most abundant genotype, G2/G1. Colors report the ratio of the indicated statistic for the thermotolerance loci to that of the genome. *, $p < 0.05$.

A.



B.



C.

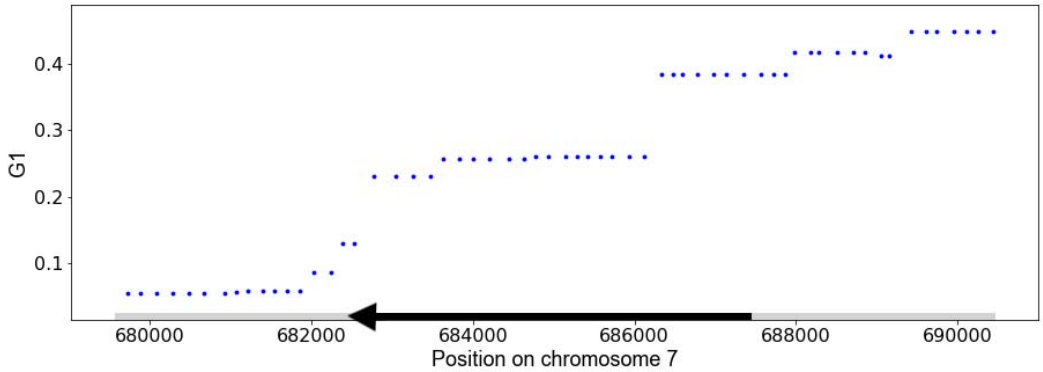


Figure 1. A peak of high allele frequency in *S. cerevisiae* populations at the 5' end of *ESP1*. Each panel shows results of analysis of allele frequency at the thermotolerance gene *ESP1* in a population of *S. cerevisiae* from (Peter et al., 2018). In each panel, the *y*-axis reports the fraction of isolates in the population that harbor the most abundant multi-locus genotype (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**.**

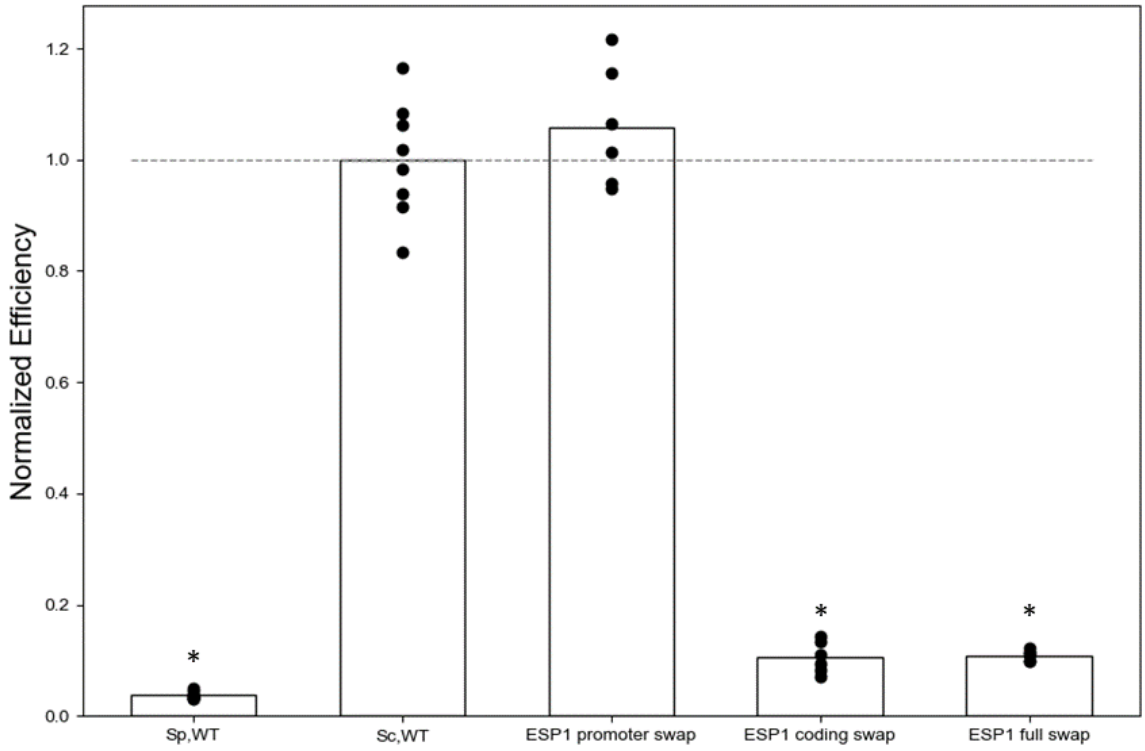


Figure 2. The *S. cerevisiae ESP1* coding region, but not the promoter, is required for thermotolerance. Each column represents results from biomass accumulation assays of a wild-type or transgenic yeast strain cultured at high temperature. The *y*-axis reports the optical density of a culture of the 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 = 6-12). 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 $p < 0.02$ in a comparison against wild-type *S. cerevisiae*. Culture data at 28°C are given in Figure S2.

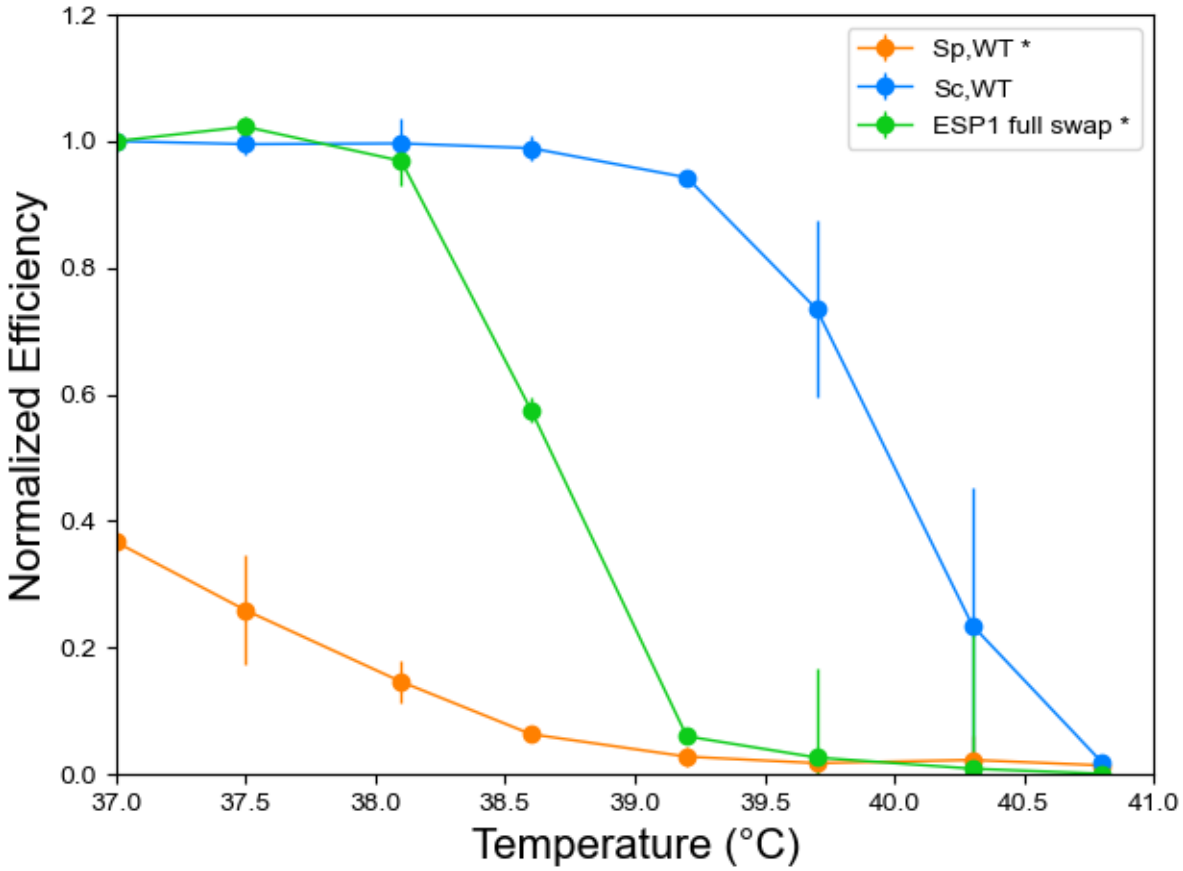
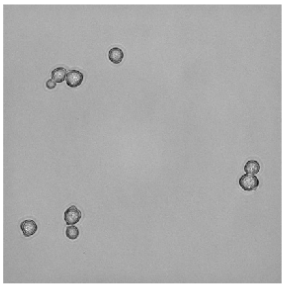
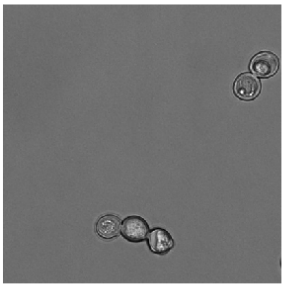
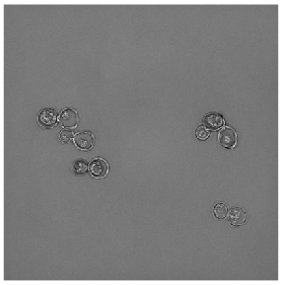
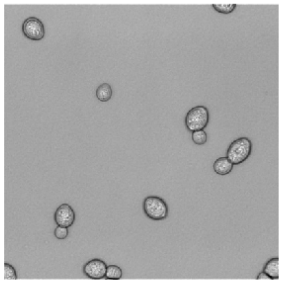
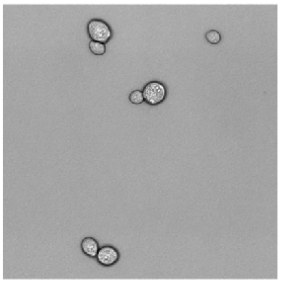
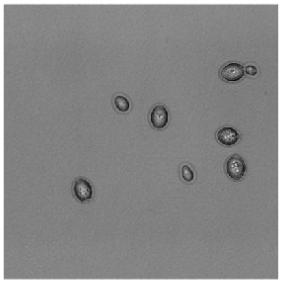


Figure 3. Growth function of *S. paradoxus ESP1* declines sharply with temperature. Each trace reports results from biomass accumulation assays of a wild-type or transgenic yeast strain across temperatures. Strain labels are as in Figure 2. The *y*-axis reports the optical density of a culture of the 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.

Temperature	Sc, WT	ESP1 full swap	Sp, WT
39ºC			
28ºC			

B.

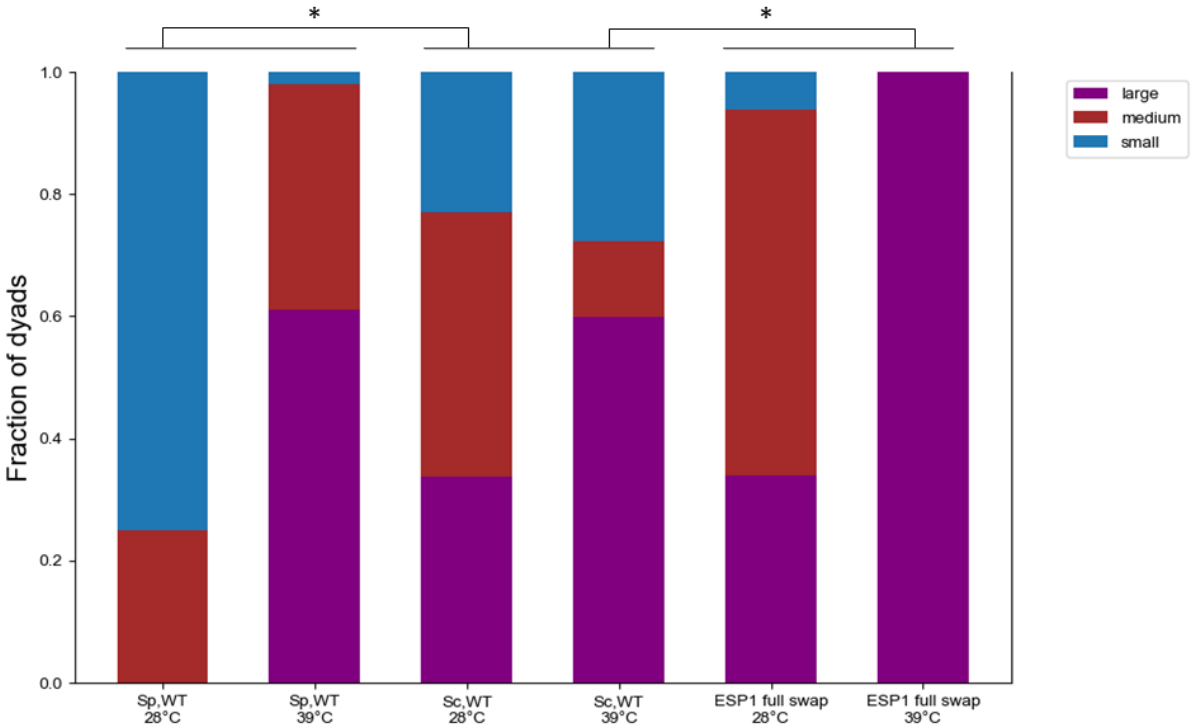


Figure 4. The *S. paradoxus* allele of *ESP1* compromises cell division at high temperature. **A**, Each panel reports a representative image of a wild-type or transgenic yeast strain after incubation for 24h at the indicated temperature. Strain labels are as in Figure 2. **B**, Each bar reports quantification 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*.