Population and comparative genetics of thermotolerance divergence between yeast species

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

*Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA; [†]Graduate Program in the Biology of Aging, University of Southern California, Los Angeles, CA; ^Department of Biology, Stanford University, Palo Alto, CA; [‡]Buck Institute for Research on Aging, Novato, CA

[1]These authors contributed equally to this work.

[2]Current address: Guardant Health, Inc., Redwood City, CA

Thermotolerance loci across yeasts

Keywords: Evolution, Genetics, Saccharomyces, Thermotolerance, Adaptation, Ancient

Corresponding Author:

Rachel Brem, Ph.D.

Department of Plant and Microbial Biology

University of California, Berkeley

312F Innovative Genomics Institute Building

2151 Berkeley Way

Berkeley, CA 94730

rbrem@berkeley.edu

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., 2018). 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., 2018) 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., 2018).

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. cerevisiae* 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 $p = 0.19$).

Polymorphism in Wine/European *S. cerevisiae*

For polymoprhism calculations in Figure S1, from 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. We tabulated the average π across a 1200 SNP window centered at each SNP, with invariant sites contributing $\pi = 0$.

***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., 2018). 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., 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 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., 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 $<e_{Scer,37}>$ 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 $e_{s,T}/<e_{Scer,37}>$. We carried out two days' worth of replicate growth experiments, and used the complete cohort of measurements of $e_{s,T}/<e_{Scer,37}>$ 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. The range and mean number of dyads scored per image and per strain are reported in Table S7.

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.

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., 2018). 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., 2018). 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 and Figure S1). As expected, the latter corresponded to troughs of polymorphism across the population as measured by the number of pairwise differences between strains (Figure S1). 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., 2018). 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., 2018), 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., 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, 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., 2018). 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.

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. BioRxiv 2021.03.18.436093.

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 (New York, N.Y.) *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.

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.

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., de Almeida, J.M.G.C.F., and Sampaio, J.P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric Saccharomyces Species. PLoS One *6*.

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 (New York, N.Y.) *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., De Chiara, M., 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. Applied and Environmental Microbiology *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 in Genetics *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*.

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 Research *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 selection. Nature Genetics *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. Nat Genet *50*, 1501–1504.

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.

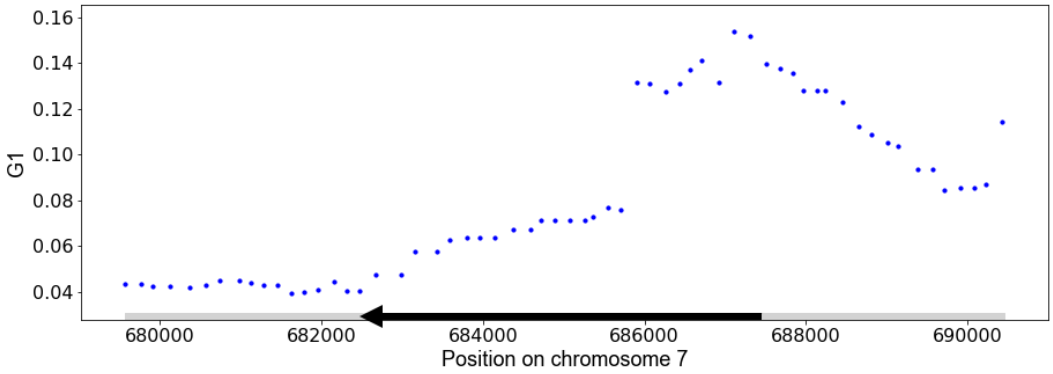
Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

ACKNOWLEDGEMENTS AND FUNDING

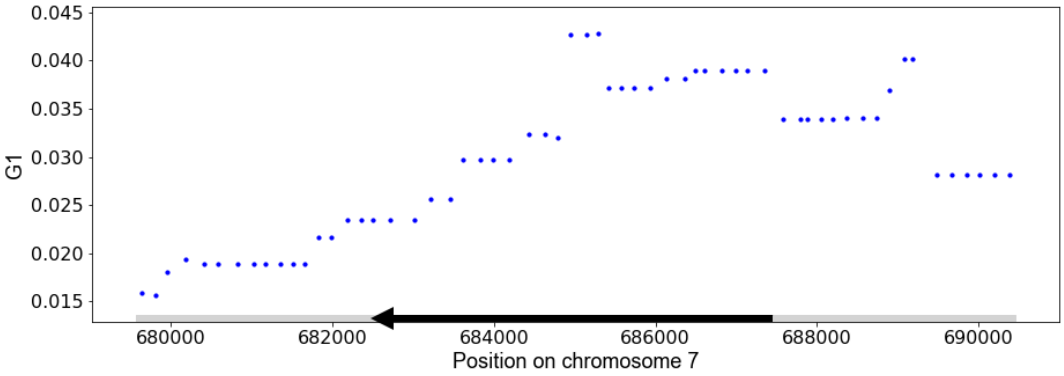
The authors thank Julie Chuong for verifying and maintaining transgenic strains, David Savage for his generosity with lab facilities and resources, Josh Schraiber for discussions, and Jeremy Roop for critical reading of the manuscript. This work was supported by NSF GRFP DGE 1752814 to M.A. and NIH R01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

FIGURES

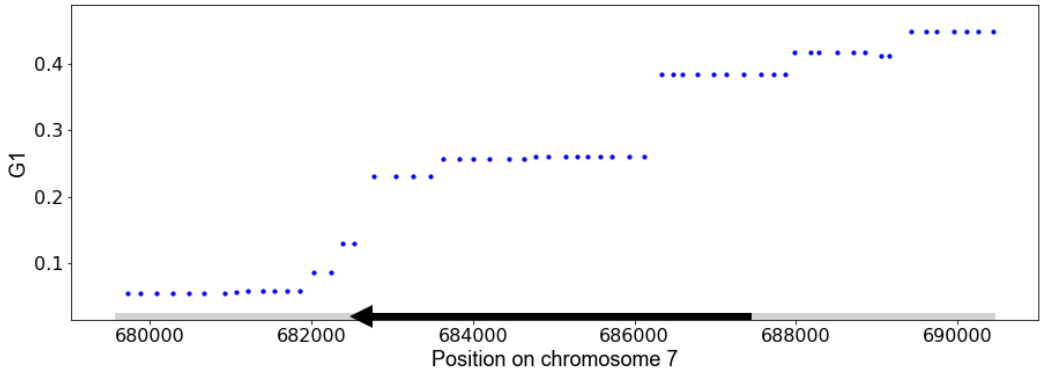
A.


0.16
0.14
0.12
0.10
0.08
0.06
0.04
G1
680000
682000
684000
686000
688000
690000
Position on chromosome 7

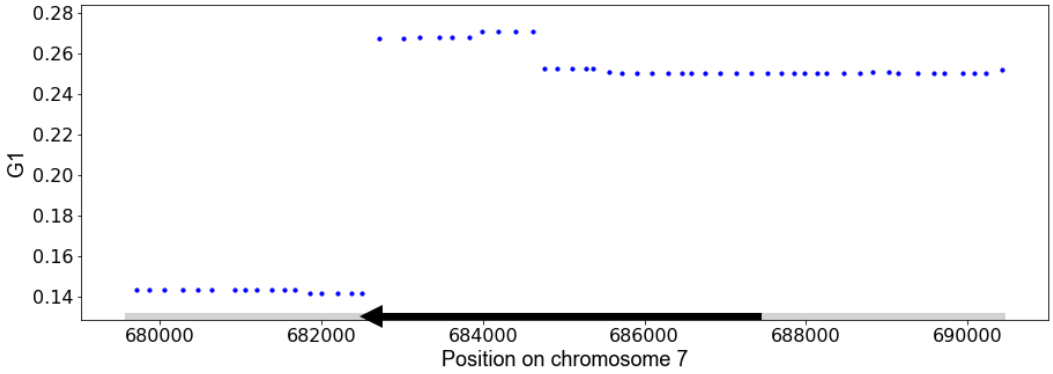
B.


0.045
0.040
0.035
0.030
0.025
0.020
0.015
G1
680000
682000
684000
686000
688000
690000
Position on chromosome 7

C.


0.4
0.3
0.2
0.1
G1
680000
682000
684000
686000
688000
690000
Position on chromosome 7

D.



E.

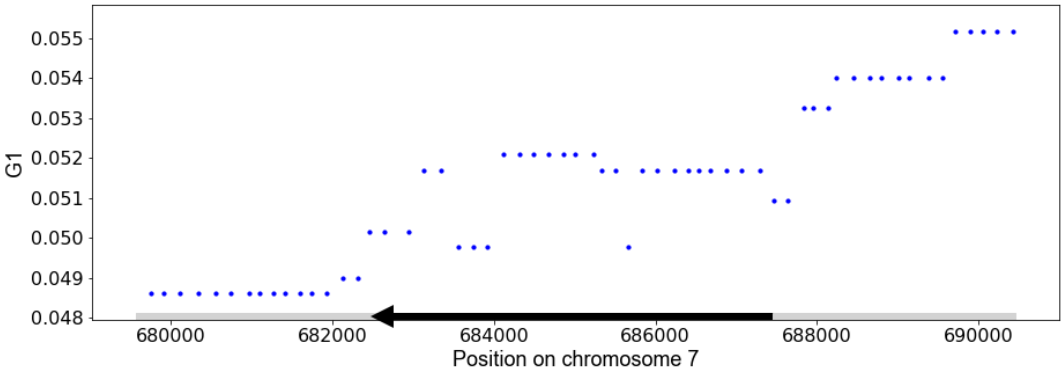


Figure 1. A peak of high allele frequency in *S. cerevisiae* populations at the 5' end of *ESP1*. Each panel shows results of analysis of allele frequency at the thermotolerance gene *ESP1* in a population of *S. cerevisiae* from (Peter et al., 2018). In each panel, the *y*-axis reports genotype homozygosity,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

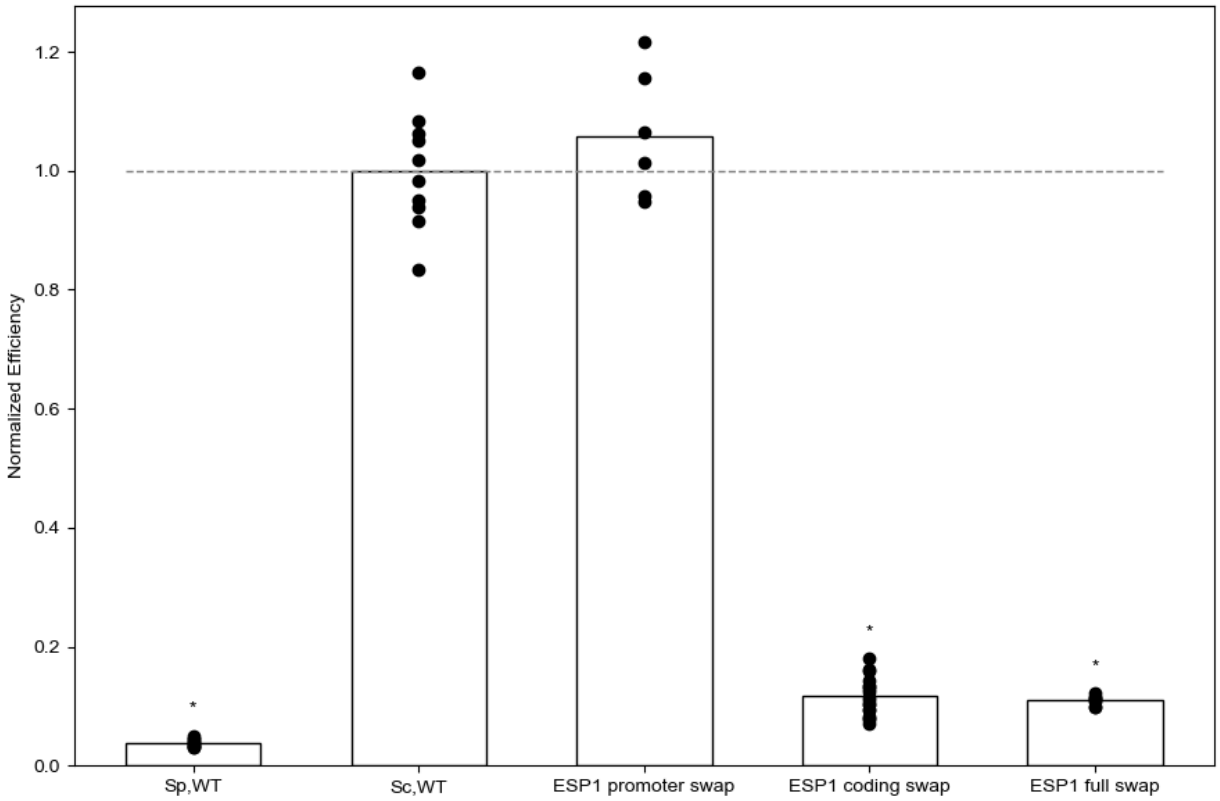


Figure 2. The *S. cerevisiae ESP1* coding region, but not the promoter, is required for thermotolerance. Each column represents results from biomass accumulation assays of a wild-type or transgenic yeast strain cultured at high temperature. The *y*-axis reports the optical density of a culture of the 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.

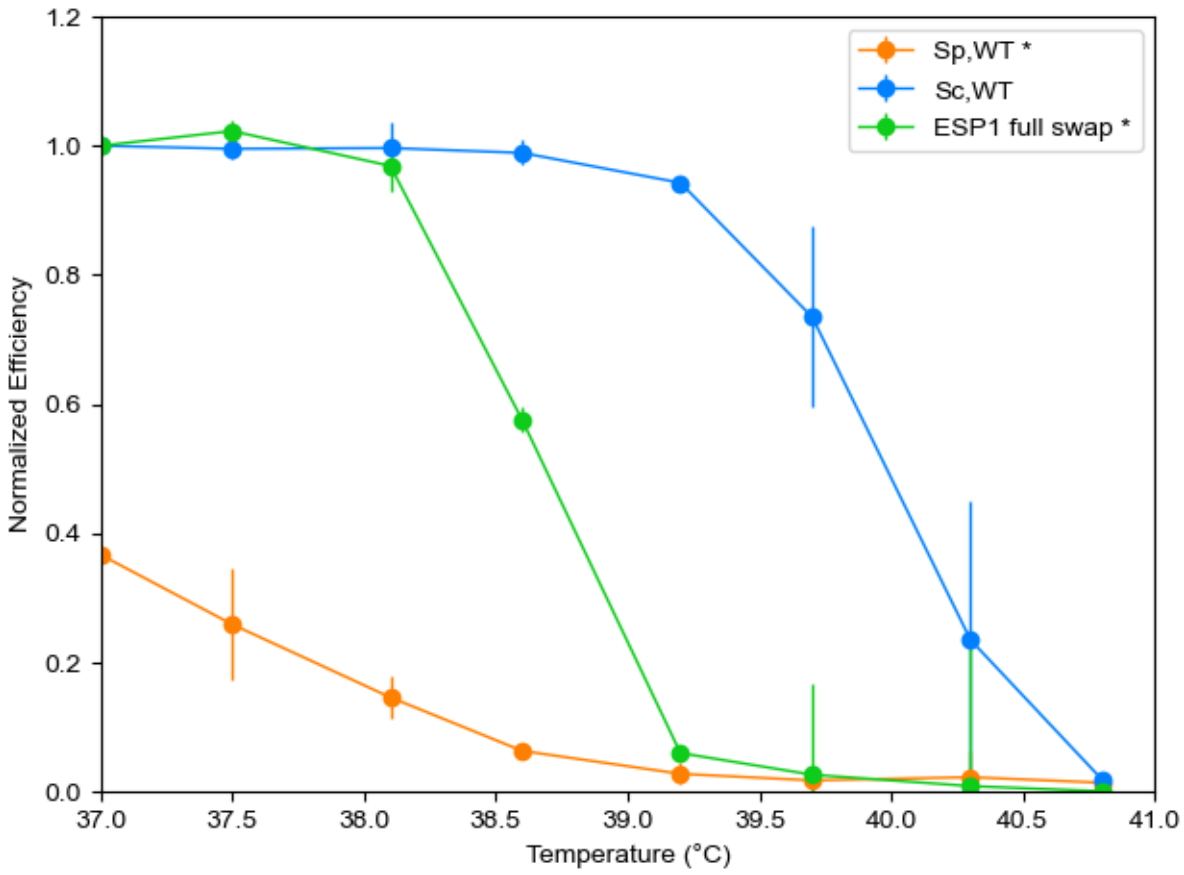
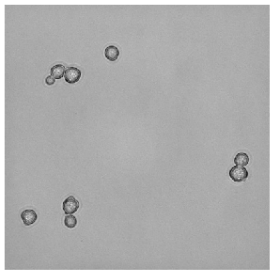
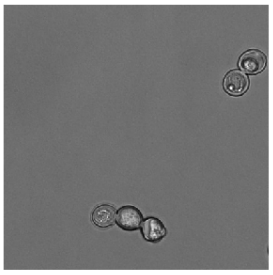
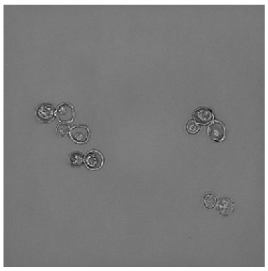
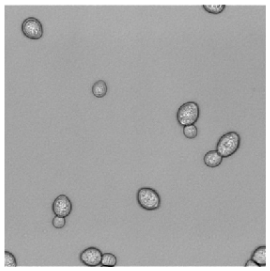
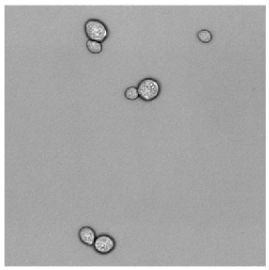
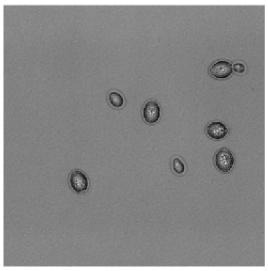


Figure 3. Growth function of *S. paradoxus ESP1* declines sharply with temperature. Each trace reports results from biomass accumulation assays of a wild-type or transgenic yeast strain across temperatures. Strain labels are as in Figure 2. The *y*-axis reports the optical density of a culture of the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

Temperature	Sc, WT	ESP1 full swap	Sp, WT
39°C			
28°C			

B.

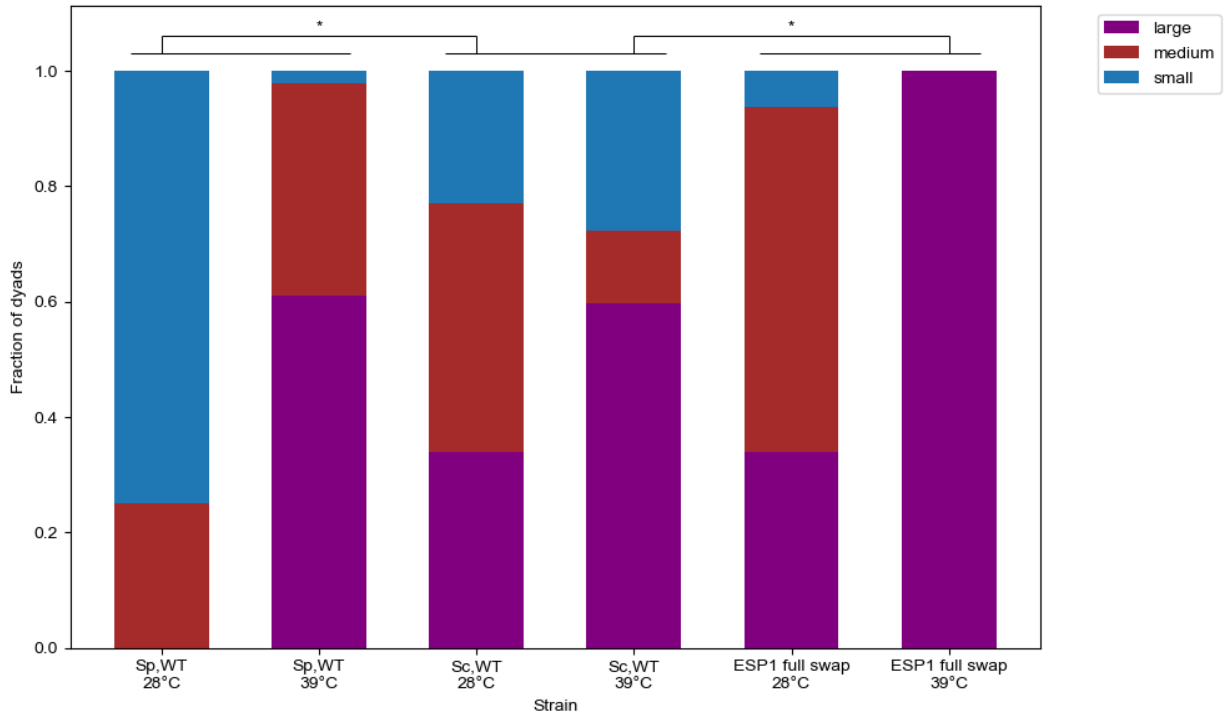


Figure 4. The *S. paradoxus* allele of *ESP1* compromises cell division at high temperature. A, Each panel reports a representative image of a wild-type or transgenic yeast strain after incubation for 24h at the indicated temperature. Strain labels are as in Figure 2. **B**, Each bar reports quantification 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*. Experiment details are given in Table S7.

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. *, $p < 0.05$.