Barcoded RH-seq illuminates the complex genetic basis of yeast thermotolerance

Melanie B. Abrams* [1], Julie N. Chuong* † [1], Faisal AlZaben*[2], Claire A. Dubin*[2], Jeffrey M. Skerker^[3], and Rachel B. Brem* ‡

*Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA; †Ph.D. Program in Biology, New York University, New York, NY; ^ Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; ‡Buck Institute for Research on Aging, Novato, CA

[1]These authors contributed equally to this work.
[2]These authors contributed equally to the work
[3]Current address: Zymergen, Inc., 5980 Horton St #105, Emeryville, CA 94608

Barcoded RH-seq and yeast thermotolerance

Keywords: Evolution, Genetics, Saccharomyces, Thermotolerance, Adaptation

Corresponding authors:
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

Jeffrey Skerker, Ph.D.
Zymergen, Inc.
5980 Horton St #105
Emeryville, CA 94608
skerker1@gmail.com

ABSTRACT

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

INTRODUCTION

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

The budding yeast *Saccharomyces cerevisiae* grows better at high temperature than any other species in its clade (Sweeney et al. 2004; Gonçalves et al. 2011; Salvadó et al. 2011; Hittinger 2013; Weiss et al. 2018). This derived and putatively adaptive trait serves as a model for the genetic study of deep evolutionary divergences. Thermosensitivity, the putatively ancestral phenotype in the clade, is borne out in *S. paradoxus*, a close sister species to *S. cerevisiae*, making the former a useful point of comparison. Our group previously used this system as a testbed to develop RH-seq, a genomic version of the reciprocal hemizygosity test (Stern 2014) that is particularly well-suited to the mapping of natural trait variation between species. We identified eight genes at which species divergence contributed to thermotolerance (Weiss et al. 2018). Combining the *S. cerevisiae* alleles of all eight genes mapped to thermotolerance recapitulated <20% of the difference between the species (AlZaben et al. bioRxiv), suggesting that many of the determinants of the trait remained undetected. If so, boosting the power and throughput of genetic mapping could help meet the challenge. We thus set out to achieve a deeper exploration of the complex genetics of yeast thermotolerance, by improvements to the RH-seq method.

MATERIALS AND METHODS

Construction of a randomly barcoded piggyBac transposase pool

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

Preparation of the backbone vector

To allow the use of BbsI as the Type IIS restriction enzyme for Golden Gate cloning of barcodes into pJR487 (see below), we first removed all three BbsI cut sites from pJR487 by introducing silent mutations

that disrupted the restriction enzyme's recognition pattern. The resulting plasmid was called pCW328. We next modified pCW328 to make a Golden-Gate-ready vector, with the final identifier pJC31, by replacing transposon nucleotides with those of a stuffer at a location 70 nucleotides from the end of the right arm of the transposon (Table S1); see Supplementary Note and Figures S3-S4 for a description of this choice. The stuffer contained 2 BbsI cut sites with custom Type IIS overhang sequences from (Lee et al. 2015), and a NotI cut site in between the 2 BbsI cut sites. All cloning steps were carried out by GenScript, Inc.

Preparation of barcode oligonucleotides

To make barcodes, we acquired an oligonucleotide pool from IDT that contained random 20 bp sequences (from hand-mixed random nucleotides) flanked by universal priming regions, U1 and U2 (Wetmore et al. 2015, Coradetti et al. 2018). These custom oligos were produced and PAGE purified by IDT. Additionally, we designed forward (FW_BbsI_JC) and reverse (REV_BbsI_JC) primers which each contained a BbsI cut site, BbsI overhang sequences complementary to the backbone vector, and either universal priming sequence (Table S2) (Coradetti et al. 2018). We set up 50 µL amplification PCR reactions with 1 µL random 20 bp barcodes as template, from a 2.5 µM stock, and 0.25 µL of each of the forward and reverse primers from a 100 µM stock. Amplification used Phusion High Fidelity polymerase (NEB) and the following cycling protocol: 98°C for 30 seconds, (98°C for 10 seconds, 58°C for 30 seconds, 72°C for 60 seconds) × 6, 72°C for 5 minutes. PCR products were purified (Zymo DNA Clean & Concentrator kit) and then combined. This yielded the final donor barcodes: random 20bp barcodes flanked by universal priming regions, with BbsI cut sites at the extreme edges.

Cloning barcodes into plasmids

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

The first reaction contained 2:1 molar ratio of vector to barcodes (4 µg of pJC31 and 128 ng of donor barcodes), 5µL of 10X T4 Ligase Buffer (ThermoFisher), 2.5 µL of T4 Ligase (ThermoFisher), 2.5µL FastDigest BpiI (ThermoFisher), and sterile water up to 50 µL. The cycling program was: 37°C for 5min, (37°C for 2min, 16C for 5min) x25, 65°C for 10 minutes. Then a mixture containing 5 µL 10X FastDigest Buffer (ThermoFisher), 3.13 µL BSA 2 mg/mL (NEB), 12.5 µL FastDigest NotI (ThermoFisher), and 12.5µL FastDigest BpiI (ThermoFisher) was spiked into the reaction and incubated at 37°C for 16 hours to digest unbarcoded backbone vectors. Ten of these reactions were combined, purified, and eluted in H_2O (Zymo DNA Clean & Concentrator). To spot-check this cloning, 5 µL of this product was transformed into 25 µL of *E. coli* 10beta electrocompetent cells (NEB). Sanger sequences across the barcode regions of 20 individually miniprepped *E. coli* colonies showed 95% barcoding efficiency.

The second reaction contained 2:1 molar ratio of vector to donor barcodes (4µg of pJC31 and 128 ng of donor barcodes), 5 µL of 10X T4 Buffer (ThermoFisher), 2.5 µL T4 Ligase (ThermoFisher), 2.5 µL BpiI (ThermoFisher), and sterile water up to 50 µL. The cycling program was: 37°C for 5min, (37°C for 2 minutes, 16°C for 5 minutes) x 25, 65°C for 10 minutes. Then a mixture containing 2.5µL 10X FastDigest Buffer (ThermoFisher), 2.5 µL G Buffer, (ThermoFisher), 3.13 µL BSA 2 mg/mL (NEB), 12.5 µL FastDigest NotI (ThermoFisher), and 12.5 µL BpiI (ThermoFisher) was spiked in the reaction and incubated at 37°C for 16 hours to digest remaining unbarcoded backbone vectors. 6 of these reactions were combined, purified, and eluted in H_2O (Zymo DNA Clean & Concentrator). Then every 5 µL of cleaned eluted product was redigested with 5µL of NotI-HF (NEB), 5 µL 10X CutSmart buffer (NEB), and 35 µL H_2O at 37°C for 16 hours then 80C for 20 minutes. The reactions were purified again (Zymo DNA Clean & Concentrator) and pooled. Spot checks of this cloning reaction proceeded as above, and Sanger sequences across the barcode regions of 20 individually miniprepped *E. coli* colonies showed 95% barcoding efficiency.

Purified plasmids from the two reactions were combined in a master tube of DNA before transforming into electrocompetent *E. coli* cells (NEB) to generate the final barcoded piggyBac pool (final identifier P58). Each electroporation cuvette (BTX) contained 25 µL of 10beta electrocompetent cells (NEB) and 5 µL of cleaned master tube DNA from the previous golden gate barcoding step. We performed 21 electroporation reactions in total using the Bio-Rad GenePulser Xcell machine set to 2.0 kV, 200 Ohms,

25 µF. After electroporation, each culture was recovered in provided outgrowth media (NEB) by shaking at 37°C at 250 rpm for 1.5 hours. After recovery, all independent 21 electroporation reactions were combined.

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

Sequencing verification of barcoded piggyBac pool plasmid DNA for barcode diversity

To verify barcode diversity in the barcoded piggyBac plasmid pool (P58), we sequenced barcodes as follows. One frozen aliquot of P58 was inoculated into 1.25 L of LB containing carbenicillin 100 µg/mL and grown for 16 hours 37°C, 250 rpm or until it reached an OD_{600} of 2.1. This culture was gigaprepped on using a column kit (Invitrogen) to generate 5 mg of plasmid. We used this as input into a PCR with primers (Table S2) annealing to the universal priming regions flanking the barcode. These primers were dual-indexed, although in this work we only carried out sequencing of the resulting amplicon from one end (see below), such that only one index was used. The generic form of the forward primer was AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT(N1-4)xxxxxxGTCGACCTGCAGCGTACG, where the N1-4 represent variable amounts of random bases from 1-4 to help samples cluster on the Illumina lane and the (x6) represent a unique 6-bp index sequence for multiplexing samples. The generic reverse primer was CAAGCAGAAGACGGCATACGAGATxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT . Four PCR reactions used 50 ng of prepped P58 plasmid template each. Amplification used Q5 High Fidelity Polymerase (NEB) and a cycling program 98°C for 4 minutes, (98°C 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds) x25, 72°C for 5 minutes. Each PCR product was purified on a column (Zymo DNA Clean & Concentrator-5 Kit) and eluted in 10 µL prewarmed 65°C provided elution buffer (Zymo). Six µL of each were then combined and sequenced off the U2 region via Illumina amplicon sequencing, on one lane of HiSeq4000 SR50 at the Genomics Sequencing Laboratory at UC Berkeley. Reads sequenced per library are reported in Table S3. Sequencing of the *E. coli* vector pool p58 revealed 27,538,142 barcodes with an estimated sequencing error rate of 1.38% analyzed as described (Coradetti et al. 2018).

Yeast hemizygote pool construction via barcoded transposon mutagenesis

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

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

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

To generate subpools consisting of a single 50-mL culture, one colony of CW27 was inoculated into 5 mL of YPD and incubated at 28°C 200 rpm. 24h later, the OD_{600} of the overnight culture was 3.86. It was backdiluted to an OD_{600} of 0.1 in 50 mL of YPD in a 250 mL Erlenmeyer flask and grown with shaking at 28°C, 200 rpm for 5.5h. After 5.5 hours, it had reached OD_{600} ~0.9 and cells were at mid-log phase. This 50 mL culture was gently pelleted at 1000xg for 3 minutes. The pellet was washed with 25 mL sterile water and then 5 mL of 0.1 M lithium acetate (Sigma) mixed with 1X Tris-EDTA buffer (10 mM Tris-HCl

and 1.0 mM EDTA); after spin-down, to the tube was added a solution of 0.269 mg of P58 mixed 5:1 by volume with salmon sperm DNA (Invitrogen), followed by 3 mL of 39.52% polyethylene glycol, 0.12 M lithium acetate and 1.2X Tris-EDTA buffer (12 mM Tris-HCl and 1.2 mM EDTA). The tube was rested for 10 minutes at room temperature, then heat-shocked in a water bath at 37°C for 26 minutes. The tube was gently spun at 1000g for 3 minutes after which supernatant was removed. We transferred the cells to a flask and added YPD to attain an OD_{600} of ~0.35–4 in ~70 mL. Each such culture was recovered by shaking at 28°C and 200 rpm for 2 h. G418 (Geneticin; Gibco) was added to each at a concentration of 300 µg/mL to select for those cells that had taken up the plasmid, and cultures were incubated with 200 rpm shaking at 28°C for 2 days until each reached an OD_{600}/mL of ~2.5. We transferred cells from this culture, and YPD + G418 (300 µg/mL), to new 250 mL flasks at the volumes required to attain an OD_{600}/mL of 0.2 in 50 mL each. We cultured each flask with 200 rpm. shaking at 28 °C overnight until each reached an OD_{600}/mL of 3.43. To cure transformants of the P58 URA3+ plasmid, we spun down 10% of this master culture, and resuspended in water with the volume required to attain a cell density of 1.85 OD_{600}/mL. Four mL of this resuspension were plated (1 mL per 24.1 cm x 24.1 cm plate) onto plates containing complete synthetic media with 5-fluorooritic acid (0.2% dropout amino acid mix without uracil or yeast nitrogen base (US Biological), 0.005% uracil (Sigma), 2% D-glucose (Sigma), 0.67% yeast nitrogen base without amino acids (Difco), 0.075% 5-fluorooritic acid (Zymo Research)). After incubation at 28°C to enable colony growth, colonies were scraped off all 4 plates and combined into water at the volume required to attain 44 OD_{600}/mL, yielding the transposon mutant hemizygote subpool. This was aliquoted into 1 mL volumes with 10% dimethylsulfoxide and frozen at −80°C.

To generate subpools consisting of five 50-mL cultures, one colony of CW27 was inoculated to 100 mL of YPD in a 250 mL Erlenmeyer flask and incubated shaking at 28°C, 200 rpm. Twenty-four hours later, the OD_{600}/mL of the overnight culture was OD_{600}/mL 3.89. The overnight culture was backdiluted to OD_{600}/mL 0.1 in 250 mL of YPD and incubated for 5.5 hours at 28°C, 200 rpm. After 5.5 hours, the OD_{600}/mL reached 0.9 and cells were split into five 50 mL conical tubes, and subjected each to heat shock as above. We then transferred all cells from this post-transformation culture to one 1L flask and added fresh YPD to attain OD_{600}/mL 0.4 in ~750 mL YPD. The transformed culture was recovered by shaking at 28°C, 200rpm, for 2 hours. G418 (300mg/ul) was added to select for the transposed cells. The culture continued shaking for 48 hours or until the OD_{600}/mL reached 2.1. This culture was then backdiluted to create a new culture at OD_{600}/mL 0.2 in 500 mL of YPD with 300mg/µL G418 shaking for 24h 28°C, 200 rpm until OD_{600}/mL ~3.4 The curing, scraping, and freezing steps were the same as above.

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

Tn-seq mapping of yeast hemizygote pool

Tn-seq library preparation

To associate barcoded transposon insertions to genomic location in the hemizygote pool, which we refer to as Tn-seq, we first sequenced barcoded transposon insertions according to the methods of (Weiss et al. 2018) as follows. Each OD44 aliquot of each subpool or final pool was thawed on ice, and its genomic DNA (gDNA) was harvested with the ZR Fungal/Bacterial DNA MiniPrep Kit (Zymo Research). gDNA was resuspended in DNA elution buffer (Zymo Research) prewarmed to 65°C, and its concentration was quantified using a Qubit 4.0 fluorometer. Illumina transposon sequencing (Tn-seq) library construction was as described previously. Briefly, gDNA was sonicated and ligated with common adapters, and for each fragment deriving from a barcoded transposon insertion in the genome, a sequence containing a barcode, a portion of the transposon, and a portion of its genomic context (the barcoded transposon–genome junction) was amplified using one primer homologous to the U1 region immediately upstream of barcode and another primer homologous to a region in the adapter. See Table S2 for the transposon-specific primer (“forward primer”), where Ns represent random nucleotides, and the indexed adapter-specific primer (“reverse primer”), where the six xs are a unique index used for multiplexing multiple

libraries onto the same Illumina sequencing lane. Amplification used Jumpstart polymerase (Sigma) and the following cycling protocol: 94°C for 2 minutes, (94°C for 30 seconds, 65°C for 20 seconds, 72°C for 30 seconds) × 25, 72°C for 10 minutes. Sequencing of paired-end reads of 150 bp was done over two lanes on a HiSeq4000 at Novogene Corporation (Sacramento, CA) and one lane on a NovaSeq SP at the Genomics Sequencing Laboratory at UC Berkeley (Berkeley, CA). Reads sequenced per library are reported in Table S4.

Tn-seq data analysis

Tn-seq data of our hemizygote pool was analyzed with the pipeline (v1.1.4) from (Coradetti et al. 2018) (https://github.com/stcoradetti/RBseq/tree/master/Old_Versions/1.1.4), adapted to map barcoded transposon insertions with the following modifications. If a compliant barcode was found, then, instead of scanning positions for the end of the insertion from a sequence specified by a model file, we searched for the final 22 base pairs of the right arm of the piggyBac transposon allowing for two mismatches. We annotated our mapped insertions to positions in the yeast genome as follows. We converted the annotation file from (Weiss et al. 2018), for the *S.* cererevisiae D1373 x *S. paradoxus* Z1 hybrid, to a compliant GFF3 file using Another GFF Analysis Toolkit (AGAT) - Version: v0.4.0 (https://github.com/NBISweden/AGAT). Then, we used a custom Jupyter notebook to annotate the pool file generated by the RBseq mapping software. We mapped 548,129 uniquely barcoded and mapped transposon insertions in our final P75 hemizygote pool. The annotation script, GFF3 file, and modified mapping script are available at https://github.com/melanieabrams-pub/RH-seq_with_barcoding.

Barcoded RH-seq competition cultures

For the thermotolerance competition at 37°C, one aliquot of the yeast hemizygote pool was thawed and inoculated into 150 mL of YPD in a 250 mL unbaffled Erlenmeyer flask and grown for six hours at 28°C, 200 rpm. This pre-culture (T_0, at OD_{600}/mL of 1.22) was backdiluted into 12 10 mL competition cultures at 200 rpm at each of 28°C and 37°C, with a starting OD_{600}/mL of 0.02 or 0.05 in at 28°C and 37°C respectively. These competition cultures were maintained within logarithmic growth through back-dilutions into fresh tubes of 10 mL of YPD at the same optical density as the starting culture, for a total of 10-15 generations. Dilutions for the 28°C competition cultures were performed after 8.5, 18.5, and 25.5 hours after the T0 timepoint, and dilutions for the 37°C competition cultures were performed after 8.5, 18.5, 25.5 hours and 32.5 hours after the T0 timepoint. The entire cell culture was harvested from each of these biological replicate tubes for sequencing as biological replicates.

For thermotolerance competition at 36°C (Table S6), competition cultures were grown as above with the following differences. The high temperature was 36°C, instead of 37°C. The pre-culture (T_0, at OD_{600}/mL of 0.693 after 5.5 hours at 28°C, 200 rpm) was backdiluted to a starting OD_{600}/mL of 0.02 for competition cultures at 36°C. Dilutions for both the 28°C and 36°C competition cultures were performed after 8.5, 18.5 and 25.25 hours after the T0 timepoint. Eleven instead of 12 replicates were carried out at 28°C.

Barcode quantification for barcoded RH-seq competition cultures

Bar-seq library preparation
To determine the abundance of barcoded transposon mutant hemizygote clones after selection, we sequenced barcodes insertions as follows. Each cell pellet from a selection sample was thawed on ice, and its genomic DNA (gDNA) was harvested with the Zymo QuickDNA Kit (Zymo#D6005). gDNA was resuspending in DNA elution buffer (Zymo Research) prewarmed to 65°C, and its concentration was quantified using a Qubit 4.0 fluorometer. The barcode insertion was amplified as above (see *Sequencing verification of barcoded piggyBac pool plasmid DNA for barcode diversity*). Each PCR product was purified on a column (Zymo DNA Clean & Concentrator) and eluted in 10 µL prewarmed 65°C provided elution buffer (Zymo). 6 µL of each were then combined and sequenced off the U2 region by Illumina sequencing on one lane of HiSeq4000 SR50 at the QB3 Genomics Sequencing Laboratory at UC Berkeley.

Bar-seq data analysis

Bar-seq data was analyzed as follows. Barcodes from the sequencing data were counted using the RBSeq pipeline (v1.1.4) from (Coradetti et al. 2018) (https://github.com/stcoradetti/RBseq/tree/master/Old_Versions/1.1.4). The resulting poolCount file was converted into a format compatible with the RH-seq pipeline of (Weiss et al. 2018), as modified by (Abrams et al. 2021). RH-seq analysis then proceeded essentially as in (Abrams et al. 2021), with slight changes as follows. We excluded from consideration genes with fewer than three transposon insertion mutants represented in the competition sequencing data per allele. Of the retained genes, for each mutant, we tabulated the quantity $a_{experimental,i}$, the average, across technical replicates, of sequencing-based abundances measured after the competition culture in biological replicate *i* of growth at the experimental temperature (36 or 37°C), and, separately, we calculated $a_{control,i}$, the analogous quantity from growth at the control temperature (28°C), for i = [1,12]. We then took the mean of the latter and used it to tabulate the temperature effect on the mutant in replicate *i*, $t_i = \log_2(a_{experimental,i}/a_{control,mean})$. We eliminated a mutant if the coefficient of variation of this quantity exceeded 10. With the remaining mutants (Tables S5 and S6), for a given gene, we compiled the vector of the *t* measurements across replicates and all hemizygotes with each species' allele of the hybrid disrupted in turn, and we used the Mann-Whitney test to compare these two vectors, with Benjamini-Hochberg correction for multiple testing (Tables S8 and S9). For a given gene, we calculated the effect size as the difference between two values: the $\log_2$(abundance at 37°C/abundance at 28°C) of the average insert in the *S. cerevisiae* allele, and the analogous quantity among inserts in the *S. paradoxus* allele of the gene. The RBseq-compatible version of the RH-seq pipeline which accepts barcode identifiers can be accessed at https://github.com/melanieabrams-pub/RH-seq_with_barcoding.

We earmarked top candidate genes for factors contributing to *S. cerevisiae's* thermotolerance as those which had corrected Mann-Whitney $p < 0.05$, and an effect size < -0.5, *i.e.* disrupting the *S. cerevisiae* allele was associated with a strong defect in thermotolerance relative to disruption of the *S. paradoxus* allele.

Gene ontology enrichment

To test RH-seq hit genes for enrichment for overrepresentation of a particular gene ontology (GO) term, we mapped each gene to its gene ontology groups based on data from geneontology.com (Ashburner et al. 2000). We filtered out narrow GO terms with fewer than five and broad GO terms with more than 200 gene members. We also filtered out GO terms with identical membership in the genome. We excluded all genes not present in our RH-seq analysis, and thus excluded any GO terms without a member represented in the fitness data. We took the subset of the remaining GO terms with at least one member in our candidate genes. Then, we randomly sampled genes from the genome, ensuring the same proportion of essential genes as in our candidate set based on the essentiality annotations of (Winzeler 1999). We tabulated whether our random sample had greater or fewer genes with that GO term than our candidate set. We repeated this procedure 10,000 times and used the proportion of these resampled groups that had more genes the given GO term as the initial *p* value assessing the significance of the enrichment of that GO term. Then, we applied Benjamini-Hochberg correction for multiple hypothesis testing to generate final, adjusted *p*-values for the enrichment of the given GO term among our candidate genes.

To test biological process ontologies for enrichment for large magnitudes of RH-seq effect size, we filtered GO terms as above. For each retained term, we tabulated the median magnitude of the effect size of the gene members present in our RH-seq fitness data. Then, we tabulated that quantity for a random sample of the same number of genes from the genome as were present in the RH-seq data for that GO term, ensuring the same proportion of essential genes as described above. We tabulated whether this median magnitude of effect size in the random sample was greater or less than the analogous quantity of the GO term we were testing. We repeated this procedure 100 times, and used the proportion of the resampled groups that had a higher median fitness effect magnitude as the initial *p* value assessing the enrichment of large fitness effects in that GO term. For all GO terms with an initial *p* value < 0.1, we repeated this procedure 10,000 times to calculate a more precise *p* value. Then, we applied the

Benjamini-Hochberg correction for multiple hypothesis testing to generate final, adjusted *p*-values for the enrichment of the given GO term for large fitness effects.

Interaction resampling

To test RH-seq hit genes for enrichment in interactions with each other, we used the Saccharomyces Gene Database interactions, downloaded February 19, 2021. We tabulated the number of physical interactions between RH-seq hit genes, and we divided that by the total number of interactions involving one RH-seq hit gene and any other gene in the genome; call this ratio r_{true}. Then, we drew a random sample of genes from the genome, as described above for GO term resampling. We tabulated, in this random gene set, the number of physical interactions between genes in that sample, and we divided that by the total number of interactions involving one gene in the random sample and any other gene in the genome, to yield $r_{resample}$. We repeated this procedure 10,000 times, and we used the proportion of resampled groups where $r_{resample}$ was greater than or equal to r_{true} as a one-sided *p* value assessing the significance of enrichment of interactions between our genes of interest. Then, we repeated this analysis for genetic interactions, or for all interactions documented.

To construct a circos plot of interacting genes within our genes of interest, we visualized the degree of interaction according to the STRING database (Szklarczyk et al. 2021).

Molecular evolution analysis

McDonald-Kreitman analysis statistics were calculated as in (Abrams et al. 2021). Fisher's exact test was used to compute *p*-values for individual loci, and these were adjusted using the Benjamini-Hochberg correction for multiple hypothesis testing.

Branch length PAML analysis with codeML was performed as in (Dubin et al. 2020). Hits were manually inspected for the quality of the alignment, and one, *YAL026C*, was discarded for poor alignment quality leading to an artefactually high branch length.

Branch-site PAML analysis with codeML was performed as in (Abrams et al. 2021).

Data Availability Statement

Sequencing data are deposited in the Sequence Read Archive under the accession PRJNA735401. 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.

Results

Dissecting thermotolerance divergence between species by barcoded transposon mutagenesis

We reasoned that we could achieve new insights into the genetics and evolution of yeast thermotolerance by revisiting the RH-seq approach. In the reciprocal hemizygosity test, in an interspecific diploid hybrid, introducing a loss-of-function mutation in one species' allele of a given gene reveals the function of the uncovered allele from the other species. Our earlier implementation of RH-seq (Weiss et al. 2018) generated such loss-of-function mutations via untargeted transposon mutagenesis in the hybrid followed by phenotyping-by-sequencing. In a scheme of this kind, tagging the transposon with sequenceable oligonucleotide barcodes can simplify library construction and reduce cost, improving replication and precision (Wetmore et al. 2015). We established an RH-seq reagent with this feature by generating a pool of plasmids, each of which encoded a barcoded copy of the piggyBac transposon and its transposase (Figure S1). We used these plasmids to mutagenize a *S. cerevisiae* x *S. paradoxus* hybrid (the F1 from a mating of strains DVBPG1373 and Z1 respectively) (Figure S1). We catalogued transposon insertions in the resulting pool of hemizygotes, by culturing the pool in standard conditions and using the DNA as input into a sequencing library primed with a common site on the transposon to sequence into the genome (Figure 1A).

We then used our barcoded hemizygote pool for RH-seq mapping of genes underlying the divergence in thermotolerance between *S. cerevisiae* and *S. paradoxus*. For this, we subjected the hemizygotes to highly replicated growth assays at high temperature (37°C), alongside controls at a permissive temperature (28°C). We then used DNA from each culture as input into barcode sequencing to quantify thermotolerance of each hemizygote in turn (Figure 1B). We established a genomic analysis pipeline for reciprocal hemizygosity tests, including quality control measures and multiple testing correction (see Methods). The results revealed 83 genes at which variation impacted high-temperature growth to a degree significant at a 5% false discovery rate (Figure 2 and Table S7). These RH-seq hits represented a much more extensive mapped set than that reported from our non-barcoded approach (Weiss et al. 2018), reflecting the power afforded by the highly-replicated barcoded method to detect even quite small effects.

In the results, as a positive control we first inspected the set of genes known to contribute to thermotolerance divergence from our earlier study (*AFG2, APC1, CEP3, DYN1, ESP1, MYO1, SCC2*, and *DYN1*) (Weiss et al. 2018). Several did not meet the experiment-wide statistical thresholds we had set for RH-seq hits (Figure S2), suggesting an appreciable false negative rate overall. However, manual inspection made clear that hemizygosity effects at all gold-standard thermotolerance loci were borne out: in each case, strains with disruptions in the *S. cerevisiae* allele, and a wild-type copy of the *S. paradoxus* allele, had worse thermotolerance than did strains with only the *S. cerevisiae* allele intact (Figure S2). Likewise, our RH-seq hit list also included *HFA1* (Figure 2B and Table S7) which was reported and validated separately as a determinant of thermotolerance differences between yeast species (Li et al. 2019). On the strength of these controls, we considered our deep sampling of thermotolerance loci to serve as a useful proof of concept for the improved RH-seq method.

Functional-genomic analysis of thermotolerance genes

We considered that downstream analysis would be most informative with a focus on the strongest and most evolutionarily relevant sources of mapping signal. In light of our interest in explaining the thermotolerance of purebred *S. cerevisiae*, we earmarked the 44 genes at which the *S. cerevisiae* allele boosted the trait most dramatically relative to that of *S. paradoxus* (Figure 2 and Table S7). In what follows, we refer to these genes as our top RH-seq hits, and we analyze them as our highest-confidence predictions for factors that nature would have used in evolving the *S. cerevisiae* phenotype.

We sought to use these top thermotolerance loci to explore potential functional mechanisms underlying the trait. Encoded proteins of these mapped genes, when analyzed together, had physical interactions

with each other to an extent beyond the null expectation (resampling $p = 0.014$). Dominating the latter signal were RH-seq hits annotated in functions of the cell cycle (Figure 3). Similarly, chromosome segregation and mitosis factors were enriched overall among our top hits (Figure 4A). A complementary quantitative analysis, using the entire RH-seq data set, likewise revealed striking effects of allelic variation on phenotype in one particular mitosis gene group, encoding components of the septin ring (Figure 4B). These results dovetail with previous phenotypic and genetic characterization of the trait, including the breakdown of cell division in heat-treated *S. paradoxus* (Weiss et al. 2018), and support a model in which *S. cerevisiae* acquired thermotolerance in part by resolving this cell cycle defect.

We anticipated that the genetics of thermotolerance would also involve mechanisms beside mitosis, since the survival advantage of *S. cerevisiae* at high temperature persists to some extent even in conditions when cells do not divide (AlZaben et al. bioRxiv). Consistent with this picture, functional-genomic tests revealed enrichment for secretion genes and for regulatory factors in our top RH-seq hits (Figure 4A). Annotations in transcription and translation, mitochondrial function, and signaling were also apparent in our top thermotolerance loci (Figure 2B). These trends suggest that evolution would have built the trait in *S. cerevisiae* by tweaking an array of housekeeping mechanisms, beside the aspects of the mechanism that involve the cell cycle machinery.

Evolutionary analysis of thermotolerance genes

We expected that sequence analyses of the genes we had mapped to thermotolerance by RH-seq could shed light on the evolutionary history of the trait. Toward this end, we first used a phylogenetic approach focused on the *Saccharomyces sensu stricto* clade, in which we quantified species-specific evolutionary rates of each gene in turn. The results revealed an enrichment for long branches leading to *S. cerevisiae* as an average across the top hits from thermotolerance RH-seq, with no such signal for other lineages (Figure 5A). Most striking was the accelerated evolution in *S. cerevisiae* of *TAF2* and *BUL1*, encoding a transcription initiation factor and ubiquitin ligase adaptor respectively (Figure 5B). We conclude that, as a group, our focal thermotolerance genes have evolved quickly in *S. cerevisiae*; alleles in this species and, likely, their phenotypic effects, represent a derived state in the clade.

Next, to explore the driving forces for these sequence changes, we quantified protein evolutionary rates in our top thermotolerance genes, using a branch-site phylogenetic modeling approach (Yang 2007). Emerging from the latter were codons with evidence for positive selection along the *S. cerevisiae* lineage, in the amino acid permease *GNP1*, the kinetochore DNA binding factor *CBF2*, and the sister chromatid cohesion factor *CTF18* (Figure 6). Thus protein evolution in *S. cerevisiae* has been driven by positive selection at these loci, putatively as a consequence of effects on the thermotolerance phenotype. Interestingly, McDonald-Kreitman tests (McDonald and Kreitman 1991) on population-genomic data did not detect an overall excess of amino acid variation relative to synonymous changes, at these three genes or any other RH-seq hit locus (Table S10). Together, our sequence analyses of thermotolerance loci reveal a wealth of alleles unique to *S. cerevisiae*, some with detectable signatures of selection, and suggest that at many genes, effects on phenotype could be mediated by just a few amino acid variants.

Discussion

RH-seq power and the interpretation of mapped loci

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

Our application to yeast thermotolerance serves as an informative model for the performance of RH-seq on highly genetically complex traits. We pinpointed dozens of genes at which species-level variation contributes to growth at high temperature. And yet we also observed evidence for a sizeable false negative rate among our RH-seq results, since some validated thermotolerance loci from our earlier

screen did not appear among the RH-seq hits here. Likewise, a separate RH-seq mapping of yeast species' differences in growth under milder heat stress revealed little signal above the noise (Table S6 and Table S9), likely reflecting very weak genetic effects under this condition. We thus expect that, as would be true for a classical linkage or association scan, the statistical power of an RH-seq experiment is a function of signal to-noise, genetic complexity, and genetic effect size; and that many thermotolerance loci remain to be identified even in our very deep set of screen results from high-temperature growth.

We also noted a type of biological false positive among our data—statistically significant mapped loci that are unlikely to have contributed to the thermotolerance of *S. cerevisiae*. Most salient among the latter are alleles from *S. cerevisiae* that compromised the trait instead of boosting it, when uncovered in the hybrid. Analogously, in linkage mapping results, the effect of an allele in recombinant progeny from a cross often does not conform to that expected from the respective parent's phenotype (Burke and Arnold 2001; Brem and Kruglyak 2005). Since such a discrepancy in RH-seq can arise from epistatic interactions between a locus of interest and the hybrid background, molecular validation will be necessary to confirm the role of our mapped genes in thermotolerance in the purebred species.

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

Cellular and molecular mechanisms of thermotolerance

Our top RH-seq hits revealed strong evidence for chromosome segregation and other mitosis functions as a linchpin of *S. cerevisiae* thermotolerance. As a complement to earlier characterization of six such genes (*APC1, ESP1, DYN1, MYO1, CEP3,* and *SCC2*) (Weiss et al. 2018; Abrams et al. 2021), we now report seven new thermotolerance determinants that function in cell division (*MEC1, MLH1, CTF13, CTF18, MCM21, CBF2,* and *MYO2*). The emerging picture is one in which the ancestor of modern-day *S. cerevisiae*, faced with dysfunction of a slew of mitotic factors at high temperature, acquired variants across the genome to shore up their activity under these conditions.

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

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

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

Acknowledgements

This work was supported by NSF GRFP DGE 1752814 to M.A. and NIH R01 GM120430 to R.B.B. The authors thank Morgan Price and Lori Huberman for invaluable discussions about barcoded transposon mutagenesis and sequencing, and Abel Duarte for cloning advice.

REFERENCES

Abrams MB, Dubin CA, AlZaben F, Bravo J, Joubert PM, Weiss CV, Brem RB. 2021. Population and comparative genetics of thermotolerance divergence between yeast species. G3 Genes|Genomes|Genetic. doi:https://doi.org/10.1093/g3journal/jkab139. [accessed 2021 May 14]. https://academic.oup.com/g3journal/advance-article/doi/10.1093/g3journal/jkab139/6259129.

AlZaben F, Chuong JN, Abrams MB, Brem RB. bioRxiv. Joint effects of genes underlying a temperature specialization tradeoff in yeast. Evolutionary Biology. [accessed 2021 May 14]. http://biorxiv.org/lookup/doi/10.1101/2021.03.18.436093.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. Nat Genet. 25(1):25–29. doi:10.1038/75556.

Boyle EA, Li YI, Pritchard JK. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 169(7):1177–1186. doi:10.1016/j.cell.2017.05.038.

Brem RB, Kruglyak L. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A. 102(5):1572–1577. doi:10.1073/pnas.0408709102.

Burke JM, Arnold ML. 2001. Genetics and the fitness of hybrids. Annu Rev Genet. 35:31–52. doi:10.1146/annurev.genet.35.102401.085719.

Coradetti ST, Pinel D, Geiselman GM, Ito M, Mondo SJ, Reilly MC, Cheng Y-F, Bauer S, Grigoriev IV, Gladden JM, et al. 2018. Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. Bohlmann J, editor. eLife. 7:e32110. doi:10.7554/eLife.32110.

Dubin CA, Roop JI, Brem RB. 2020. Divergence of Peroxisome Membrane Gene Sequence and Expression Between Yeast Species. G3: Genes, Genomes, Genetics. 10(6):2079–2085. doi:10.1534/g3.120.401304.

Gonçalves P, Valério E, Correia C, Almeida JMGCF de, Sampaio JP. 2011. Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric Saccharomyces Species. PLOS ONE. 6(6):e20739. doi:10.1371/journal.pone.0020739.

Hittinger CT. 2013. Saccharomyces diversity and evolution: a budding model genus. Trends Genet. 29(5):309–317. doi:10.1016/j.tig.2013.01.002.

Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. Nature. 425(6959):686–691. doi:10.1038/nature02026.

Lee ME, DeLoache WC, Cervantes B, Dueber JE. 2015. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth Biol. 4(9):975–986. doi:10.1021/sb500366v.

Leuenberger P, Ganscha S, Kahraman A, Cappelletti V, Boersema PJ, von Mering C, Claassen M, Picotti P. 2017. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science. 355(6327). doi:10.1126/science.aai7825.

Li XC, Fay JC. 2017. Cis-Regulatory Divergence in Gene Expression between Two Thermally Divergent Yeast Species. Genome Biol Evol. 9(5):1120–1129. doi:10.1093/gbe/evx072.

Li XC, Peris D, Hittinger CT, Sia EA, Fay JC. 2019. Mitochondria-encoded genes contribute to evolution of heat and cold tolerance in yeast. Sci Adv. 5(1):eaav1848. doi:10.1126/sciadv.aav1848.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 351(6328):652–654. doi:10.1038/351652a0.

Morellet N, Li X, Wieninger SA, Taylor JL, Bischerour J, Moriau S, Lescop E, Bardiaux B, Mathy N, Assrir N, et al. 2018. Sequence-specific DNA binding activity of the cross-brace zinc finger motif of the piggyBac transposase. Nucleic Acids Research. 46(5):2660–2677. doi:10.1093/nar/gky044.

Orr HA. 1998. The Population Genetics of Adaptation: The Distribution of Factors Fixed During Adaptive Evolution. Evolution. 52(4):935–949. doi:10.1111/j.1558-5646.1998.tb01823.x.

Ott J, Wang J, Leal SM. 2015. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 16(5):275–284. doi:10.1038/nrg3908.

Pierleoni A, Martelli PL, Fariselli P, Casadio R. 2007. eSLDB: eukaryotic subcellular localization database. Nucleic Acids Res. 35(Database issue):D208–D212. doi:10.1093/nar/gkl775.

Salvadó Z, Arroyo-López FN, Guillamón JM, Salazar G, Querol A, Barrio E. 2011. Temperature adaptation markedly determines evolution within the genus Saccharomyces. Appl Environ Microbiol. 77(7):2292–2302. doi:10.1128/AEM.01861-10.

Stern DL. 2014. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. Trends Genet. 30(12):547–554. doi:10.1016/j.tig.2014.09.006.

Sweeney JY, Kuehne HA, Sniegowski PD. 2004. Sympatric natural Saccharomyces cerevisiae and S. paradoxus populations have different thermal growth profiles. FEMS Yeast Res. 4(4–5):521–525. doi:10.1016/S1567-1356(03)00171-5.

Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, et al. 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research. 49(D1):D605–D612. doi:10.1093/nar/gkaa1074.

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. 2019. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 20(8):467–484. doi:10.1038/s41576-019-0127-1.

Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. Science. 324(5927):659–662. doi:10.1126/science.1169766.

Weiss CV, Brem RB. 2019. Dissecting Trait Variation across Species Barriers. Trends in Ecology & Evolution. 34(12):1131–1136. doi:10.1016/j.tree.2019.07.013.

Weiss CV, Roop JI, Hackley RK, Chuong JN, Grigoriev IV, Arkin AP, Skerker JM, Brem RB. 2018. Genetic dissection of interspecific differences in yeast thermotolerance. Nature Genetics. 50(11):1501. doi:10.1038/s41588-018-0243-4.

Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, et al. 2015. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. Moran MA, editor. mBio. 6(3):e00306-15. doi:10.1128/mBio.00306-15.

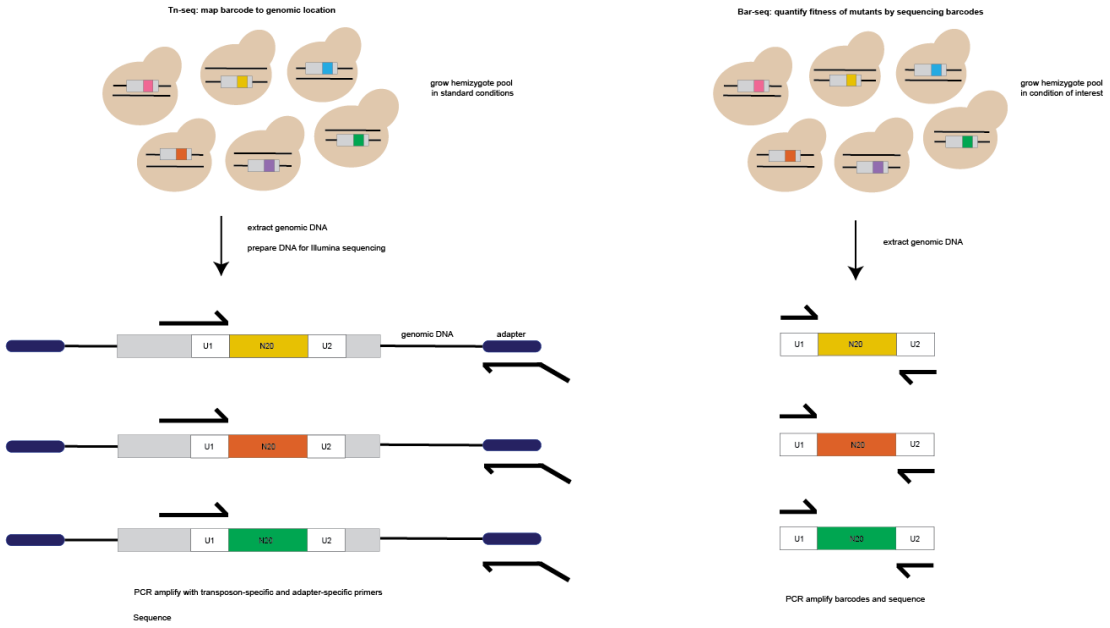
Winzeler EA. 1999. Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science. 285(5429):901–906. doi:10.1126/science.285.5429.901.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution. 24(8):1586–1591. doi:10.1093/molbev/msm088.

Yusa K, Zhou L, Li MA, Bradley A, Craig NL. 2011. A hyperactive piggyBac transposase for mammalian applications. PNAS. 108(4):1531–1536. doi:10.1073/pnas.1008322108.

FIGURES

A



B

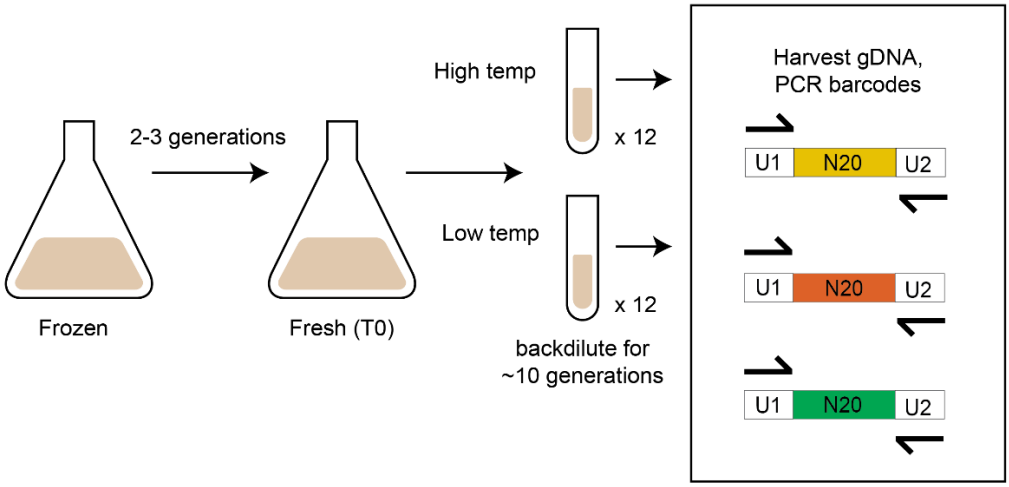
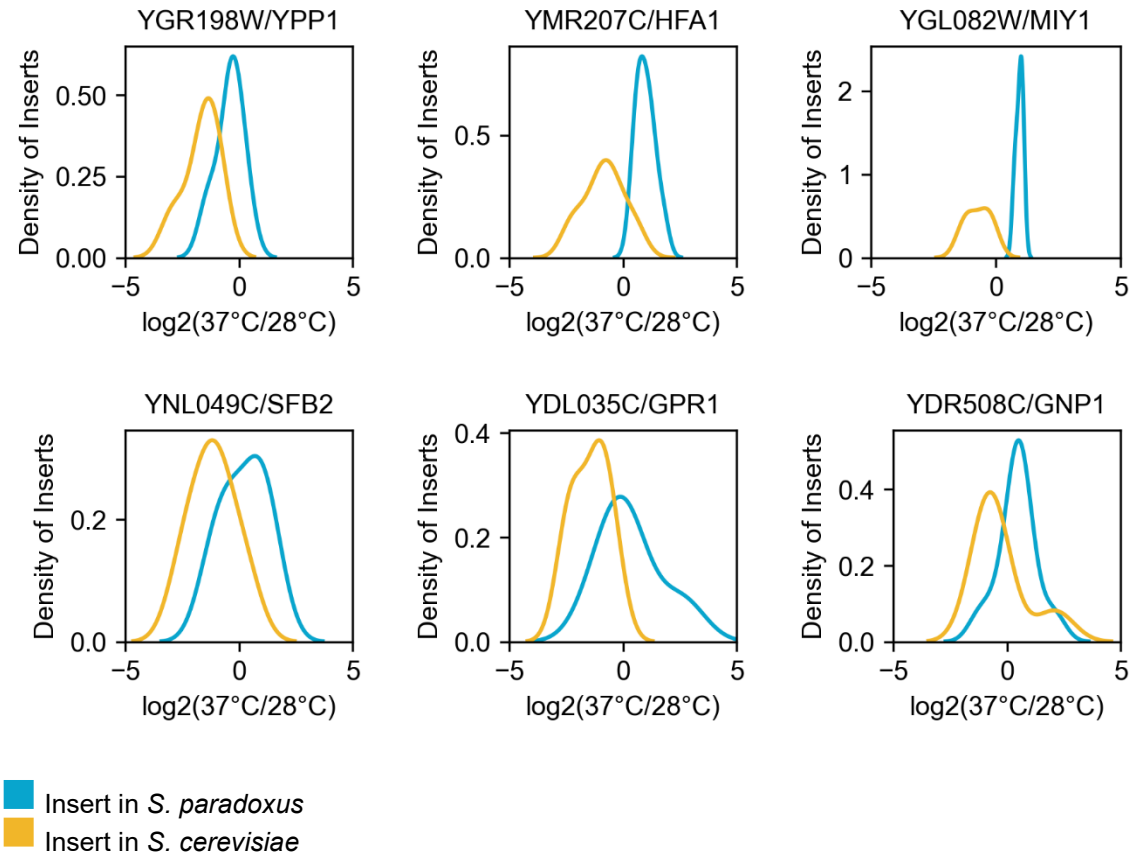


Figure 1. Barcoded RH-seq mapping of yeast thermotolerance loci. (A) Barcoded RH-seq sequencing analysis steps. Left, in a pool of *S. cerevisiae* x *S. paradoxus* hybrid hemizygotes, each harboring a transposon (grey rectangle) marked with a unique 20-mer barcode (multicolored) flanked by universal primer sites (U1 and U2), each barcode is associated with its insertion location by transposon sequencing (Tn-seq). Genomic DNA from the pool is extracted, sheared, and ligated to universal adapters (blue ovals), followed by PCR amplification with a transposon-specific primer (forward black arrow) and an adapter-specific primer (reverse black arrow) and sequencing. Right, for barcode sequencing (Bar-seq) to quantify hemizygote strain abundance after pool growth in a condition of interest, genomic DNA is used as input to PCR with primers to universal primer sites for sequencing. (B) Thermotolerance RH-seq screen design. An aliquot of the hemizygote pool was thawed and cultured in

large format, then split into small replicate cultures, each maintained at the temperature of interest by back-dilution, followed by quantification by Bar-seq.

A



B

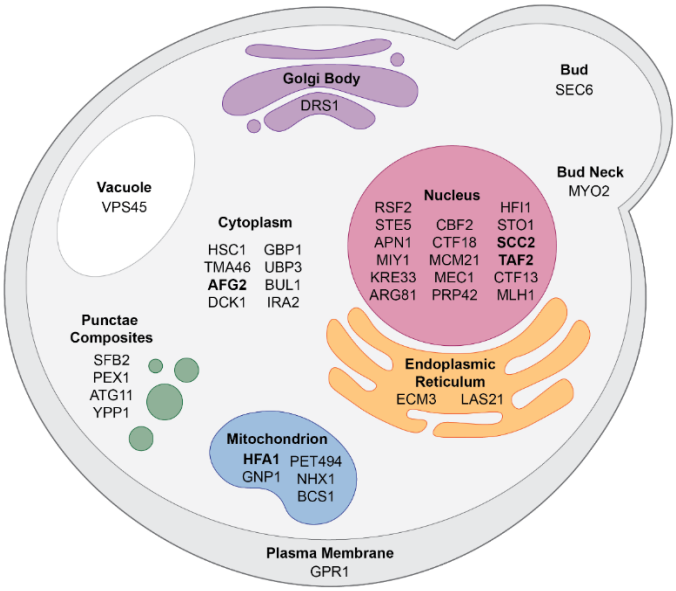


Figure 2. Hits from RH-seq mapping of yeast thermotolerance. (A) Each panel reports RH-seq results

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

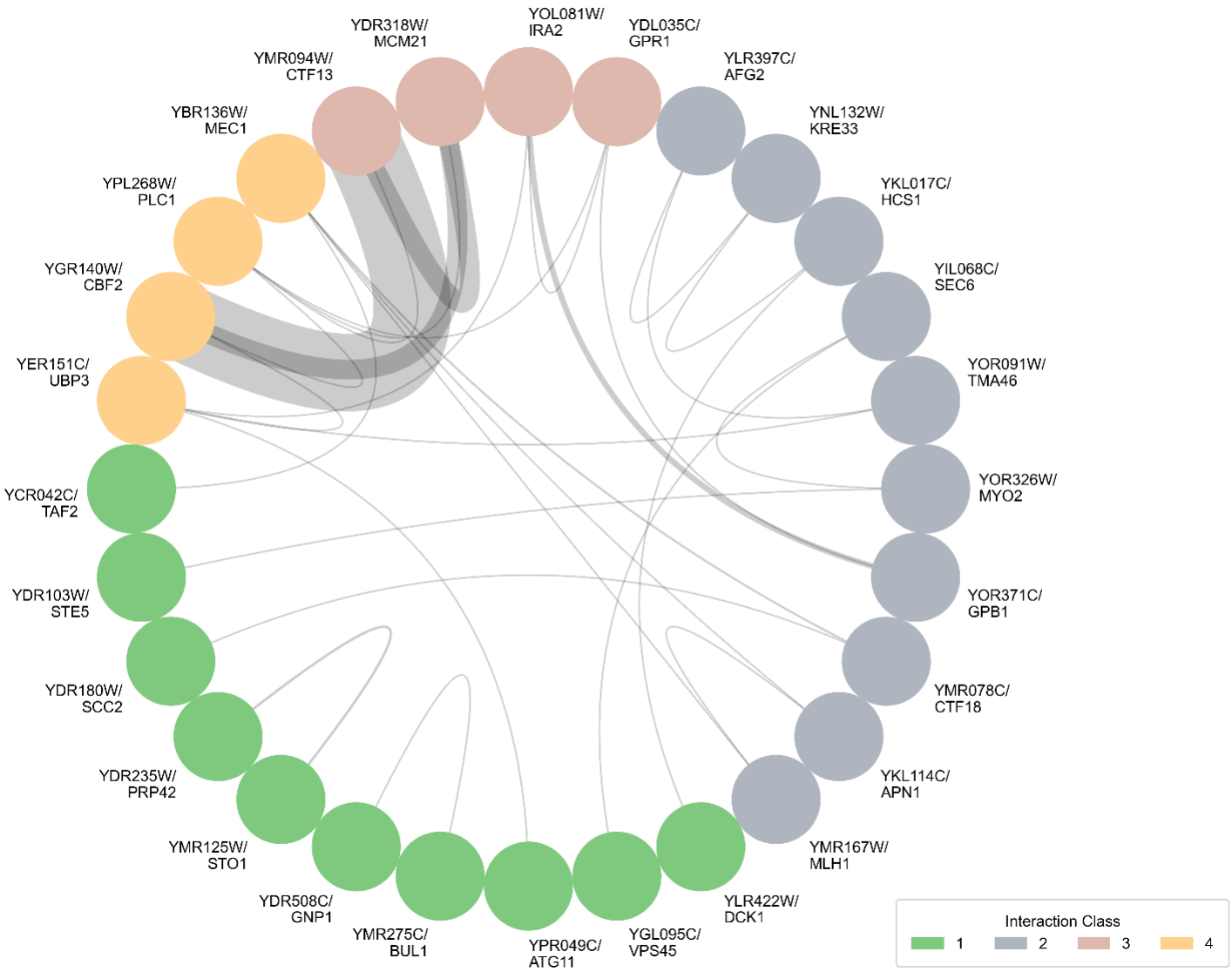


Figure 3. Interactions between thermotolerance loci. Each node represents a hit gene from RH-seq mapping of thermotolerance which interacts with at least one other such hit; each chord represents an interaction. Chords are weighted by interaction degree, and nodes are colored by degree class of their interactions.

A.

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

B.

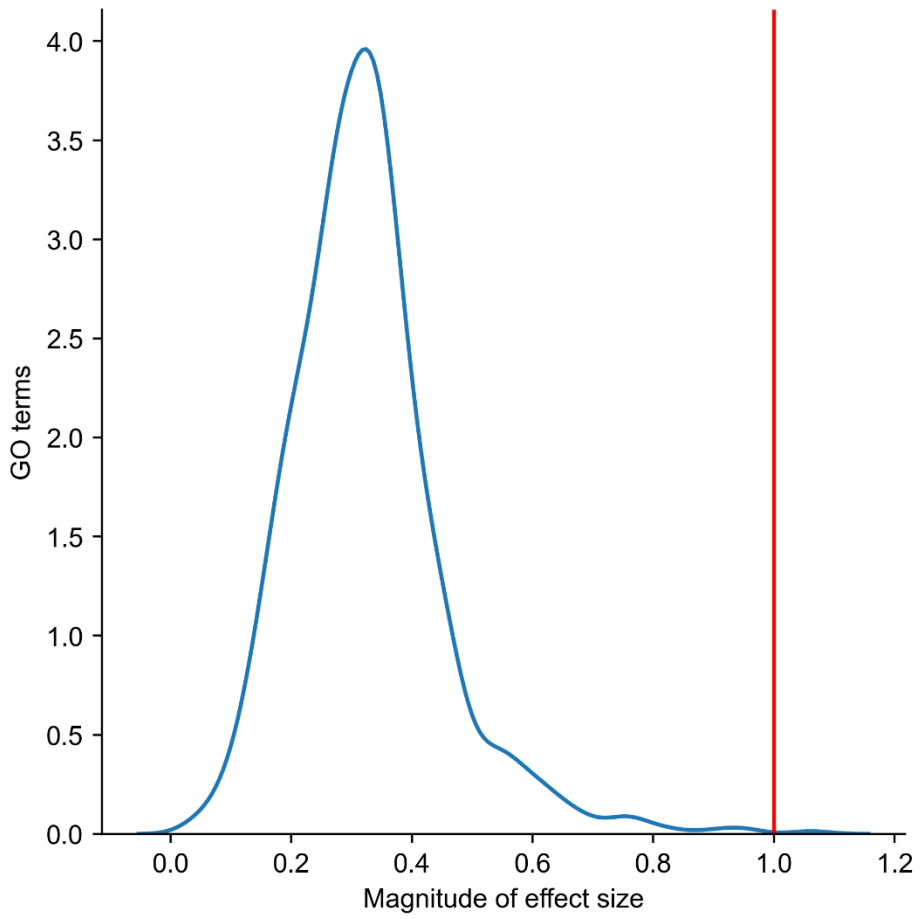
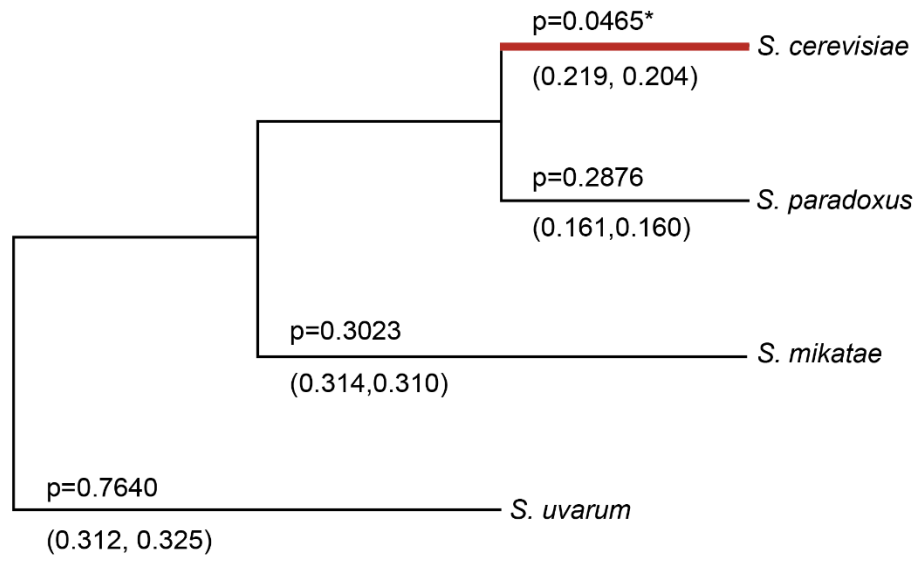


Figure 4. Functional enrichment among thermotolerance loci. (A) Each row with numerical data reports a Gene Ontology (GO) term enriched for RH-seq hit genes. $n_{observed}$, the number of RH-seq hits annotated with the term. $n_{expected}$, the number of genes annotated with the term in the same number of andomly chosen genes from the genome, as a median across samples. Adjusted *p*, resampling-based significance of the enrichment after Benjamini-Hochberg correction. Total *n*, the total number of genes annotated in the GO term in *S. cerevisiae*. (B) For each gene, we tabulated the impact of high temperature on abundance, in RH-seq data, of hemizygotes with the *S. cerevisiae* allele uncovered in the hybrid, as a difference from the analogous quantity for the *S. paradoxus* allele; the *x*-axis reports the absolute value of this effect size as a median across genes of a given GO term. The *y*-axis reports the frequency with which each effect size was detected among GO terms, given by a kernel density estimate. The vertical red line reports the effect size of the GO term septin ring assembly (GO:0000921), which is elevated to an extent unlikely under the genomic null (resampling $p < 1 \times 10^{-4}$).

A.


p=0.0465*
S. cerevisiae
(0.219, 0.204)
p=0.2876
S. paradoxus
(0.161,0.160)
p=0.3023
S. mikatae
(0.314,0.310)
p=0.7640
S. uvarum
(0.312, 0.325)

B.

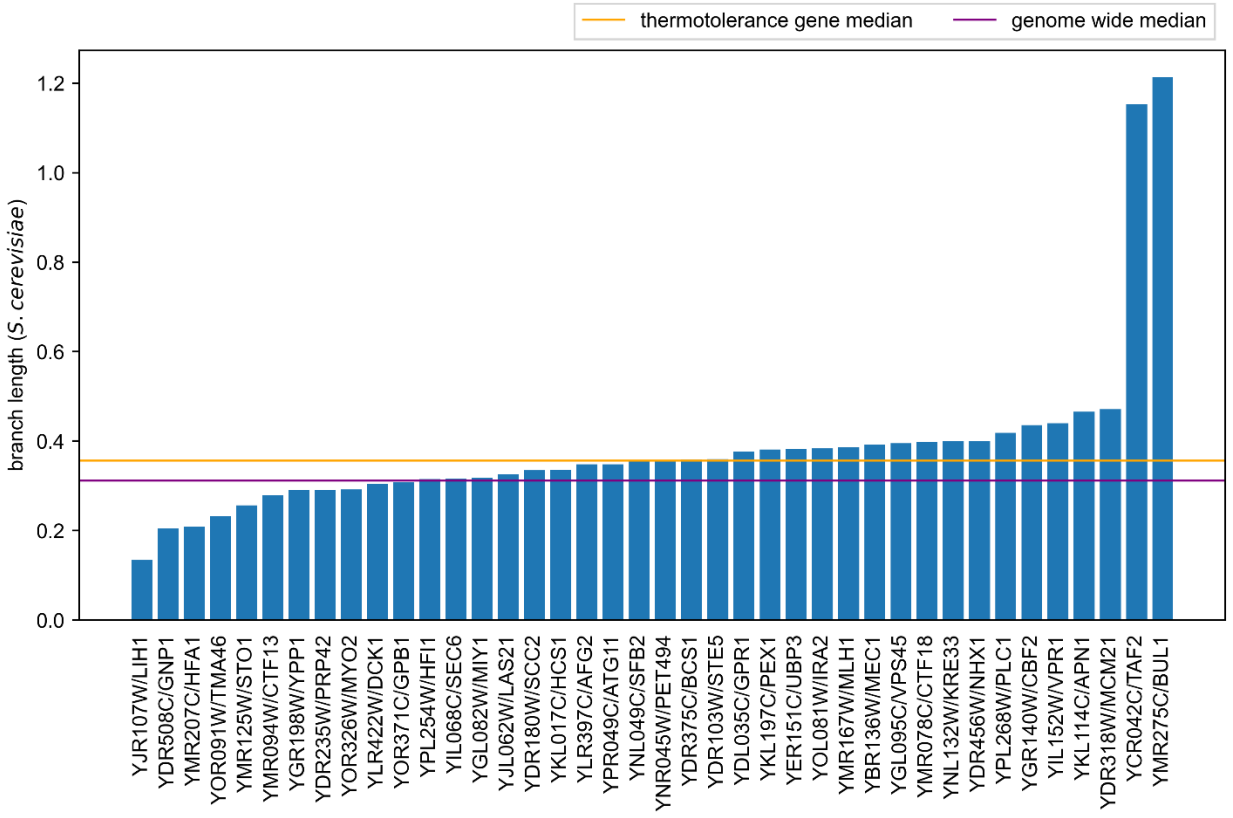

thermotolerance gene median
genome wide median
branch length (S. cerevisiae)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
YJR107W/LIH1
YDR508C/GNP1
YMR207C/HFA1
YOR091W/TMA46
YMR125W/STO1
YMR094W/CTF13
YGR198W/YPP1
YDR235W/PRP42
YOR326W/MYO2
YLR422W/DCK1
YOR371C/GPB1
YPL254W/HFI1
YIL068C/SEC6
YGL082W/MIY1
YJL062W/LAS21
YDR180W/SCC2
YKL017C/HCS1
YLR397C/AFG2
YPR049C/ATG11
YNL049C/SFB2
YNR045W/PET494
YDR375C/BCS1
YDR103W/STE5
YDL035C/GPR1
YKL197C/PEX1
YER151C/UBP3
YOL081W/IRA2
YMR167W/MLH1
YBR136W/MEC1
YGL095C/VPS45
YMR078C/CTF18
YNL132W/KRE33
YDR456W/NHX1
YPL268W/PLC1
YGR140W/CBF2
YIL152W/VPR1
YKL114C/APN1
YDR318W/MCM21
YCR042C/TAF2
YMR275C/BUL1

Figure 5. **Accelerated evolution of thermotolerance loci.** Shown are results of analyses of branch length of hit genes from RH-seq mapping of thermotolerance, as inferred from gene trees and normalized for gene length. (A) The value over each branch reports the *p*-value in a resampling-based test for long branches on the indicated *Saccharomyces* lineage. The values below each branch report (median branch length of thermotolerance loci, median branch length of randomly sampled genes). The one significant lineage, that leading to *S. cerevisiae*, is marked in red. (B) Each vertical bar reports inferred branch length, along the *S. cerevisiae* lineage, of the indicated RH-seq hit gene. Horizontal lines report median branch lengths across the indicated gene sets.

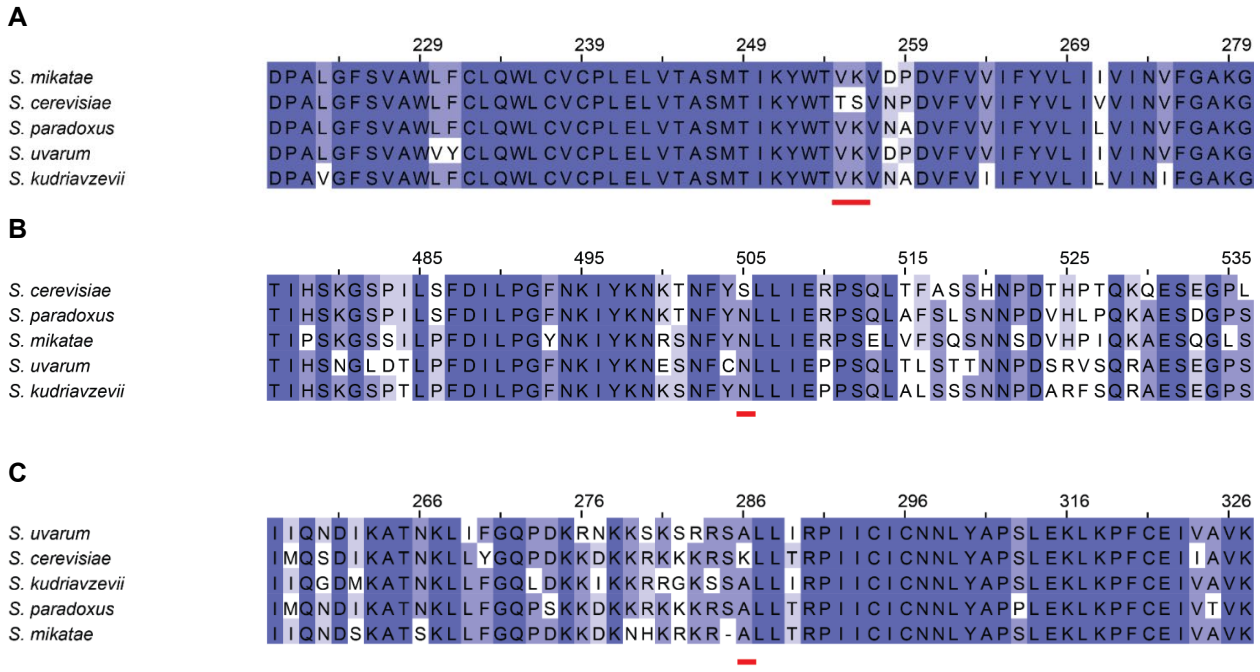
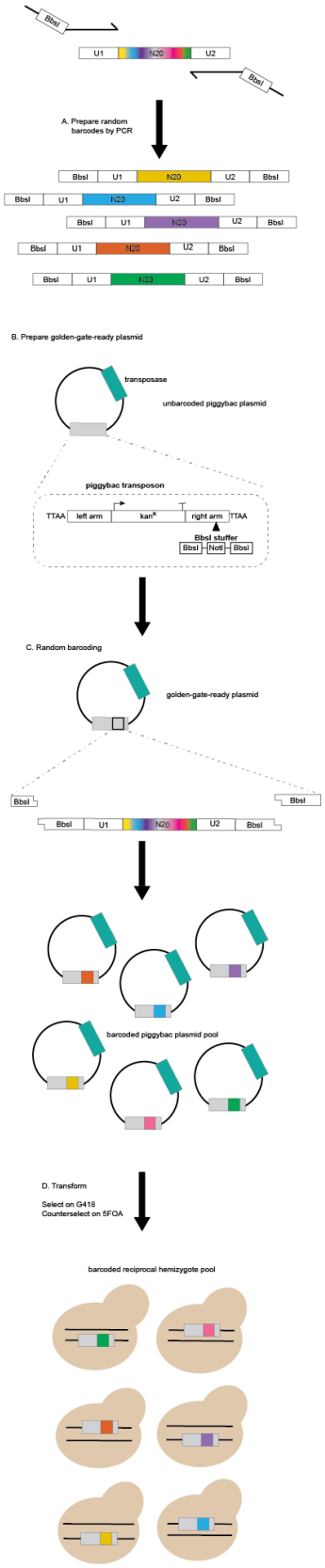
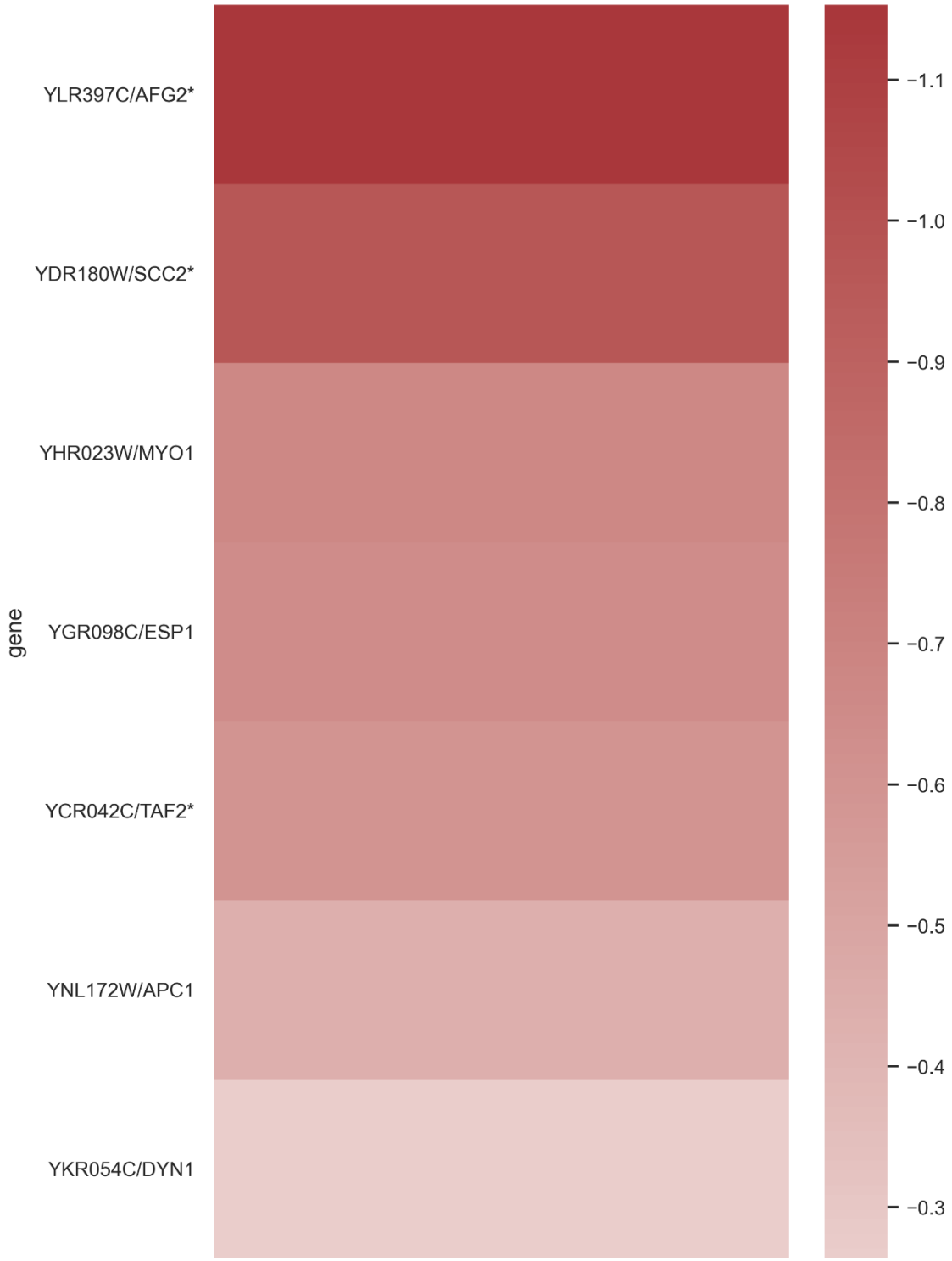


Figure 6. Codons under positive selection in thermotolerance loci. Each panel shows the amino acid sequence context of a codon (red bar) inferred to be under positive selection along the *S. cerevisiae* lineage, in a hit gene from RH-seq thermotolerance mapping. Alignments are colored by percent identity. (A) *YDR508C/GNP1*. (B) *YGR140W/CBF2*. (C) *YMR078C/CTF18*.

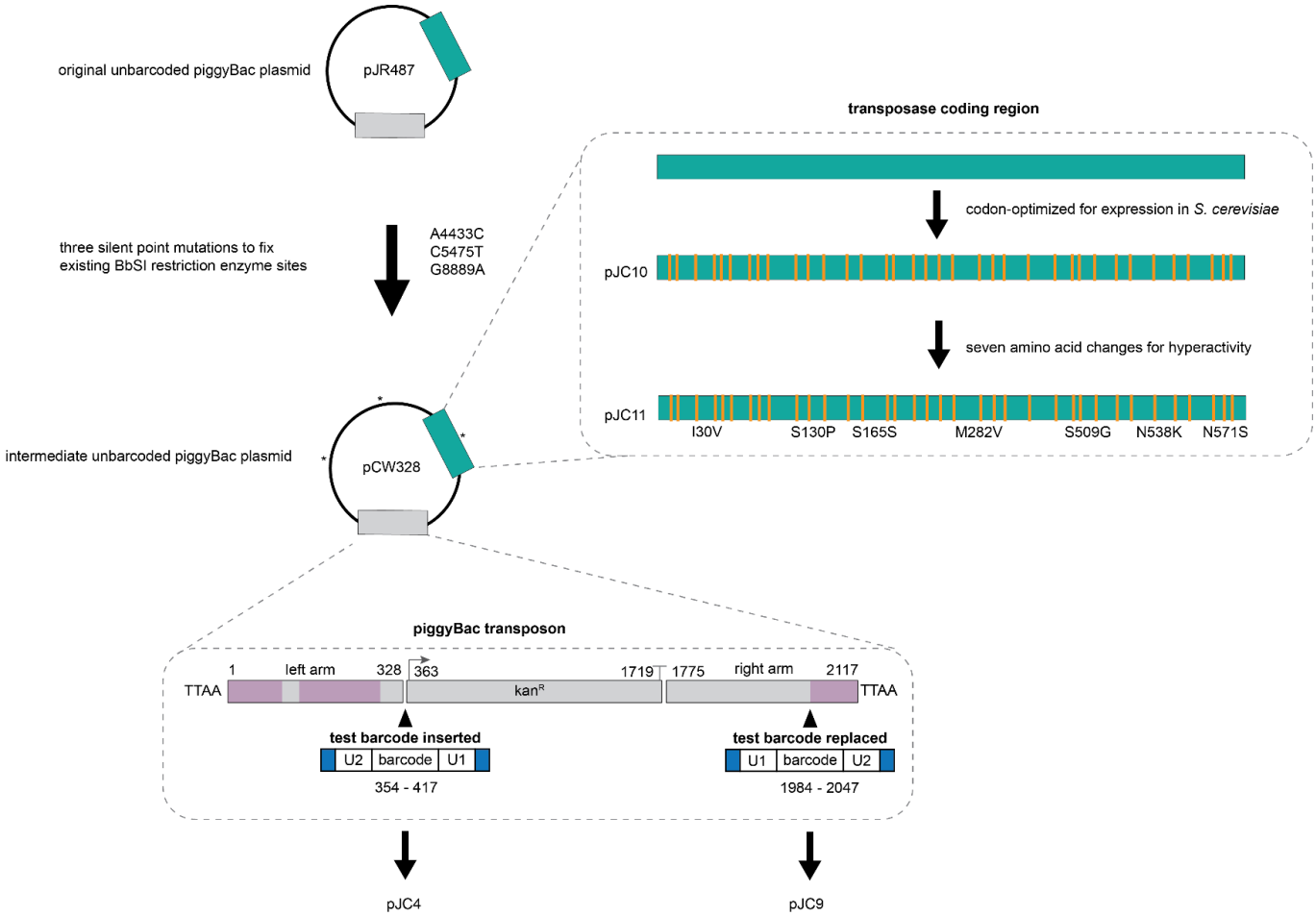

BbsI
U1
N20
U2
BbsI
A. Prepare random barcodes by PCR
BbsI U1 N20 U2 BbsI
BbsI U1 N20 U2 BbsI
BbsI U1 N20 U2 BbsI
BbsI U1 N20 U2 BbsI
BbsI U1 N20 U2 BbsI
B. Prepare golden-gate-ready plasmid
transposase
unbarcoded piggybac plasmid
piggybac transposon
TTAA
left arm
kan^R
right arm
TTAA
BbsI stuffer
BbsI
NotI
BbsI
C. Random barcoding
golden-gate-ready plasmid
BbsI
BbsI
BbsI U1 N20 U2 BbsI
barcoded piggybac plasmid pool
D. Transform
Select on G418
Counterselect on 5FOA
barcoded reciprocal hemizygote pool

Supplementary Figure 1. Making barcoded hemizygotes in a yeast hybrid background for RH-seq. (A) A pool of random N20 barcodes (colors), each flanked by universal priming sites (U1 and U2), was used as input into a PCR with primers containing recognition sites for the BbsI type IIS restriction enzyme. (B) In a plasmid harboring an un-barcoded piggyBac transposon (gray rectangle) (the kanamycin resistance cassette, kan^R, flanked by left and right transposon arms) and transposase (teal rectangle) (Weiss et al. 2018), a 42 nucleotide stuffer sequence, consisting of two BbsI restriction enzyme sites flanking a NotI restriction enzyme site and custom overhang sequences (Lee et al. 2015), replaced 42 nucleotides in the right arm of the transposon. (C) BbsI digestion of the barcodes and stuffer-containing plasmid, followed by stuffer loss and ligation of a barcode into each plasmid, yielded a pool of barcoded plasmids. (D) Transformation of the barcoded transposase plasmid into *S. cerevisiae* x *S. paradoxus* hybrids, followed by transposition and plasmid loss, yielded a pool of marked transposon hemizygote mutants in the hybrid background.

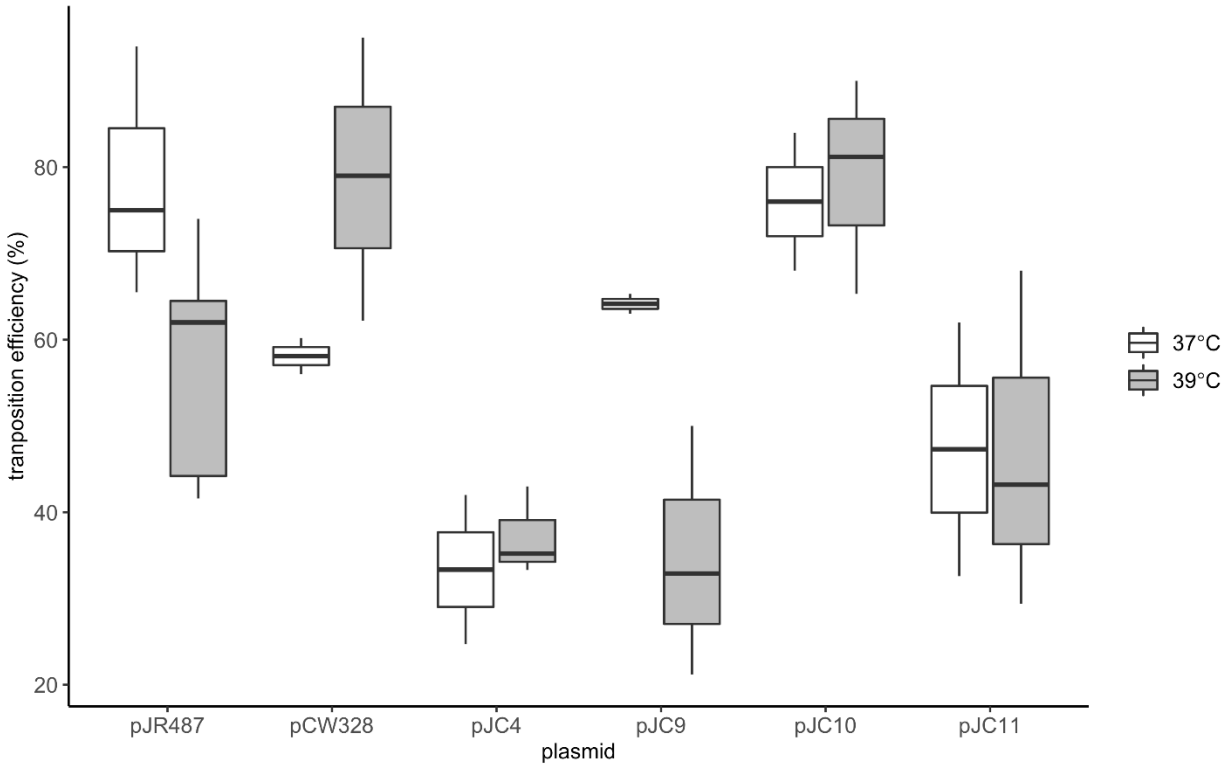


Supplementary Figure 2. Impact on high-temperature growth of allelic variation at genes from a previous thermotolerance screen. Each row reports the abundance, in RH-seq data, of hemizygotes with the *S. cerevisiae* allele of a given gene uncovered by a transposon insertion in the hybrid, relative to the analogous quantity for the *S. paradoxus* allele, of a gene at which allelic variation was previously

reported to impact thermotolerance (Weiss et al. 2018). Genes marked with asterisks were significant at $p < 0.05$, after quality control for noise and number of inserts and multiple-hypothesis correction.



Supplementary Figure 3. Modifying yeast piggyBac to test barcode insertion positions and transposase optimization. Left, a plasmid from (Weiss et al. 2018) containing the unbarcoded piggyBac transposon (gray) and transposase (teal) was modified to eliminate three BbsI restriction enzyme sites, and used as a backbone for further modifications. Right, test plasmids were mutated at transposase sites designed to optimize codons and increase activity (Yusa et al. 2011). Bottom, test plasmids were modified to incorporate into the transposon a single 20 nucleotide barcode flanked by universal priming regions and custom two-nucleotide overhang sequences (blue squares), either by insertion between the 3'-most end of the left arm and 5' end of the TEF promoter of the kanamycin cassette (bottom left) or replacing endogenous nucleotides inside the right arm of the transposon (bottom right). Pink rectangles indicate transposase binding sites from (Morellet et al. 2018). Each plasmid was used in tests of modification effects on transposition efficiency (see Figure S4)



Supplementary Figure 4. Effects of barcode insertion positions and transposase sequence on yeast piggyBac transposition efficiency. Each pair of boxes reports transposition test results from a plasmid schematized in Figure S3 in the *S. cerevisiae* x *S. paradoxus* F1 hybrid, with transformation at the indicated temperature. For a given box, the thick black line reports the median; the box extent report quartiles; whiskers report outliers.

Supplementary Table 1. Plasmids used in this study.

Oligo name	Sequence (5' to 3')	Index Identifier	Index Sequence	Notes
Barcoding cloning				
Random Barcodes U1 — N20 — U2	GATGTCCACGAGGTCTCTNNNNNNNNNNNNNNNNNNNNCGTACGCTGCAGGTCGAC			Random barcodes flanked by universal primer sites
FW_BbsI_JC	TCACACAAGTTTGTACAAAAAAGCAGGCTGGAGCTCGgaagacATCCCTGATGTCCACGAGGTCTCT			Forward primer to amplify barcodes for cloning into vector
REV_BbsI_JC	CTCAACCACTTTGTACAAGAAAGCTGGGTGGATCCgaagacCGCGTTGTCGACCTGCAGCGTACG			Reverse primer for to amplify barcodes for cloning into vector
Tn-Seq				
JC8 Tn-specific FW Primer	ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNCCCTGATGTCCACGAGGTCTCT			Forward primer to sequence transposon insertions; homologous to transposon (Figure 1A, left)
P7_MOD_TS_index1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index1	ATCACG	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index2	CGATGT	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index3	TTAGGC	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index4	TGACCA	Indexed reverse primer to sequence transposon

				insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index5	ACAGTG	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index6	GCCAAT	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index7	CAGATC	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
P7_MOD_TS_index8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Index8	ACTTGA	Indexed reverse primer to sequence transposon insertions; homologous to adapter (Figure 1A, left)
Bar-seq				
P1_BS3_IT001	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNGCACTAGTCGACCTGCAGCGTACG	IT001	ATCACG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT002	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNTGTAGCGTCGACCTGCAGCGTACG	IT002	CGATGT	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT003	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNCGGATTGTCGACCTGCAGCGTACG	IT003	TTAGGC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT004	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC	IT004	TGACCA	Indexed forward Bar-seq primer

	CGATCTNNNNACCAGTGTCGACCTG CAGCGTACG			(Figure 1A, right)
P1_BS3_IT005	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNGTGACAGTCGACCTGCAG CGTACG	IT005	ACAGTG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT006	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNNTAACCGGTCGACCTGCA GCGTACG	IT006	GCCAAT	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT007	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNNNCTAGACGTCGACCTGCA GCGTACG	IT007	CAGATC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT008	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNNNNAGTTCAGTCGACCTGC AGCGTACG	IT008	ACTTGA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT009	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNGACTAGGTCGACCTGCAG CGTACG	IT009	GATCAG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT010	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNNTTCGATGTCGACCTGCAG CGTACG	IT010	TAGCTT	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT011	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNNNCATCGGGTCGACCTGC AGCGTACG	IT011	GGCTAC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT012	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNNNNATGTTCGTCGACCTGC AGCGTACG	IT012	CTTGTA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT013	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNAACTGAGTCGACCTGCAGC GTACG	IT013	AGTCAA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT014	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC CGATCTNNCCTTGAGTCGACCTGCAG CGTACG	IT014	AGTTCC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT015	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTC	IT015	ATGTCA	Indexed forward Bar-seq primer

	CGATCTNNNACTGTAGTCGACCTGCAGCGTACG			(Figure 1A, right)
P1_BS3_IT016	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCTGCCGTCGACCTGCAGCGTACG	IT016	CCGTCC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT017	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNGAGATGGTCGACCTGCAGCGTACG	IT017	GTAGAG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT018	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNCGCCTGGTCGACCTGCAGCGTACG	IT018	GTCCGC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT019	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNAAAGTGGTCGACCTGCAGCGTACG	IT019	GTGAAA	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT020	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCGGTGGTCGACCTGCAGCGTACG	IT020	GTGGCC	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT021	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNGCTTTGGTCGACCTGCAGCGTACG	IT021	GTTTCG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT022	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNGCATGCGTCGACCTGCAGCGTACG	IT022	CGTACG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT023	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNGGTGAGGTCGACCTGCAGCGTACG	IT023	GAGTGG	Indexed forward Bar-seq primer (Figure 1A, right)
P1_BS3_IT024	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCGATGGGTCGACCTGCAGCGTACG	IT024	GGTAGC	Indexed forward Bar-seq primer (Figure 1A, right)
P2_BS3_IT001	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT001	ATCACG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT002	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGT	IT002	CGATGT	Indexed reverse Bar-seq primer

	GCTCTTCCGATCTGATGTCCACGAGGTCTCT			(Figure 1A, right)
P2_BS3_IT003	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT003	TTAGGC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT004	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT004	TGACCA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT005	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT005	ACAGTG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT006	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT006	GCCAAT	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT007	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT007	CAGATC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT008	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT008	ACTTGA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT009	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT009	GATCAG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT010	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT010	TAGCTT	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT011	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT011	GGCTAC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT012	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	IT012	CTTGTA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT013	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGT	IT013	AGTCAA	Indexed reverse Bar-seq primer

	GCTCTTCCGATCTGATGTCCACGAGG TCTCT			(Figure 1A, right)
P2_BS3_IT014	CAAGCAGAAGACGGCATACGAGATG GAACTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGATGTCCACGAGG TCTCT	IT014	AGTTCC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT015	CAAGCAGAAGACGGCATACGAGATT GACATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGATGTCCACGAGG TCTCT	IT015	ATGTCA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT016	CAAGCAGAAGACGGCATACGAGATG GACGGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCTGATGTCCACGAG GTCTCT	IT016	CCGTCC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT017	CAAGCAGAAGACGGCATACGAGATC TCTACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGATGTCCACGAGG TCTCT	IT017	GTAGAG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT018	CAAGCAGAAGACGGCATACGAGATG CGGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCTGATGTCCACGAG GTCTCT	IT018	GTCCGC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT019	CAAGCAGAAGACGGCATACGAGATTT TCACGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCTGATGTCCACGAGGT CTCT	IT019	GTGAAA	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT020	CAAGCAGAAGACGGCATACGAGATG GCCACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCTGATGTCCACGAG GTCTCT	IT020	GTGGCC	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT021	CAAGCAGAAGACGGCATACGAGATC GAAACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGATGTCCACGAGG TCTCT	IT021	GTTTCG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT022	CAAGCAGAAGACGGCATACGAGATC GTACGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCTGATGTCCACGAG GTCTCT	IT022	CGTACG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT023	CAAGCAGAAGACGGCATACGAGATC CACTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGATGTCCACGAGG TCTCT	IT023	GAGTGG	Indexed reverse Bar-seq primer (Figure 1A, right)
P2_BS3_IT024	CAAGCAGAAGACGGCATACGAGATG CTACCGTGACTGGAGTTCAGACGTGT	IT024	GGTAGC	Indexed reverse Bar-seq primer

	GCTCTTCCGATCTGATGTCCACGAGGTCTCT			(Figure 1A, right)

Supplementary Table 2. Oligonucleotides used in this study.

Experiment	Library	Reads
	RBJC009_IT013	93,277,156
	RBJC009_IT014	73,914,499
	RBJC009_IT015	79,273,122
	RBJC009_IT016	81,404,878
	RBJC45_IT013	100,794,463
	RBJC45_IT014	92,410,111
	RBJC45_IT015	91,931,624
	RBJC45_I0T16	96,631,037
	RBJC009_IT013	93,277,156
	RBJC009_IT014	73,914,499
	RBJC009_IT015	79,273,122
E. coli vector pool Bar-seq	RBJC45_IT016	96,631,037
	RMBA038A_IT001	16,042,537
	RBMA038A_IT002	19,451,725
	RBMA038A_IT003	16,313,279
	RBMA038A_IT004	19,512,043
	RBMA038A_IT005	13,625,004
	RBMA038A_IT006	22,257,379
	RBMA038A_IT007	22,385,133
	RBMA038A_IT008	16,359,601
	RBMA038A_IT009	20,118,293
	RBMA038A_IT010	19,121,369
	RBMA038A_IT011	18,887,044
	RBMA038A_IT012	15,895,905
	RBMA038A_IT013	18,922,229
	RBMA038A_IT014	26,602
	RBMA038A_IT015	16,571,744
	RBMA038A_IT016	18,783,445
	RBMA038A_IT017	17,671,257
	RBMA038A_IT018	18,040,605
	RBMA038A_IT019	16,929,985
	RBMA038A_IT020	15,319,721
	RBMA038A_IT021	16,897,401
	RBMA038A_IT022	16,372,841
37°C vs. 28°C barcoded RH-seq in yeast	RBMA038A_IT023	16,679,583
	RBMA038A_IT024	13,875,240
	RMBA039A_IT001	28,004,341
	RBMA039A_IT002	19,283,097
	RBMA039A_IT003	12,427,040
	RBMA039A_IT004	13,934,236
	RBMA039A_IT005	12,068,707
	RBMA039A_IT006	60,072,532
	RBMA039A_IT008	12,500,340
36°C vs. 28°C barcoded RH-seq in yeast	RBMA039A_IT009	25,029,516
	RBMA039A_IT010	14,765,007

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

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

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

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

Supplementary Table 5. Abundance of hemizygote mutants from thermotolerance RH-seq. Each row reports the results of sequencing one transposon insertion in the *S. cerevisiae* x *S. paradoxus* diploid hybrid after selection of the barcoded transposon mutant pool, reflecting the abundance in the pool of the respective hemizygote clone harboring the insertion. Chromosome, strand, location, and gene report the fine-scale position of the insertion. Allele, the species parent's homolog in which the transposon insertion lay. Abundance, read counts of the transposon insertion sequenced after selection of the mutant pool at the indicated temperature, normalized for library size and averaged across biological and technical replicates. Transposon insertions not detected in any replicate of the indicated selection were assigned an abundance of 1. CV, coefficient of variation over biological replicates of normalized read counts after selection at the indicated temperature. Barcode, the unique barcode identifier of the transposon insertion.

Supplementary Table 6. Abundance of hemizygote mutants from RH-seq at 36°C. Data are as in Table S5, except that RH-seq was done using 36°C as the high-temperature condition.

Hit	Description
YGR198W/YPP1	Cargo-transport protein involved in endocytosis; interacts with phosphatidylinositol-4-kinase Stt4; GFP-fusion protein localizes to the cytoplasm; YGR198W is an essential gene
YMR207C/HFA1	Mitochondrial acetyl-coenzyme A carboxylase, catalyzes the production of malonyl-CoA in mitochondrial fatty acid biosynthesis
YGL082W/YGL082W	Putative protein of unknown function; predicted prenylation/proteolysis target of Afc1p and Rce1p; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; YGL082W is not an essential gene
YNL049C/SFB2	Component of the Sec23p-Sfb2p heterodimer of the COPII vesicle coat, required for cargo selection during vesicle formation in ER to Golgi transport; homologous to Sec24p and Sfb3p
YDL035C/GPR1	Plasma membrane G protein coupled receptor (GPCR) that interacts with the heterotrimeric G protein alpha subunit, Gpa2p, and with Plc1p; sensor that integrates nutritional signals with the modulation of cell fate via PKA and cAMP synthesis
YDR508C/GNP1	High-affinity glutamine permease, also transports Leu, Ser, Thr, Cys, Met and Asn; expression is fully dependent on Grr1p and modulated by the Ssy1p-Ptr3p-Ssy5p (SPS) sensor of extracellular amino acids
YBR136W/MEC1	Genome integrity checkpoint protein and PI kinase superfamily member; signal transducer required for cell cycle arrest and transcriptional responses prompted by damaged or unreplicated DNA; monitors and participates in meiotic recombination
YML099C/ARG81	Zinc-finger transcription factor of the Zn(2)-Cys(6) binuclear cluster domain type, involved in the regulation of arginine-responsive genes; acts with Arg80p and Arg82p
YPL254W/HFI1	Adaptor protein required for structural integrity of the SAGA complex, a histone acetyltransferase-coactivator complex that is involved in global regulation of gene expression through acetylation and transcription functions
YIL152W/YIL152W	Putative protein of unknown function
YKL017C/HCS1	Hexameric DNA polymerase alpha-associated DNA helicase A involved in lagging strand DNA synthesis; contains single-stranded DNA stimulated ATPase and dATPase activities; replication protein A stimulates helicase and ATPase activities
YGR140W/CBF2	Essential kinetochore protein, component of the CBF3 multisubunit complex that binds to the CDEIII region of the centromere; Cbf2p also binds to the CDEII region possibly forming a different multimeric complex, ubiquitinated in vivo
YJR127C/RSF2	Zinc-finger protein involved in transcriptional control of both nuclear and mitochondrial genes, many of which specify products required for glycerol-based growth, respiration, and other functions
YDR375C/BCS1	Mitochondrial protein of the AAA ATPase family; has ATP-dependent chaperone activity; required for assembly of Rip1p and Qcr10p into cytochrome bc(1) complex; mutations in human homolog BCS1L are linked to neonatal mitochondrial diseases
YOR091W/TMA46	Protein of unknown function that associates with translating ribosomes; interacts with GTPase Rbg1p
YLR397C/AFG2	ATPase of the CDC48/PAS1/SEC18 (AAA) family, forms a hexameric complex; is essential for pre-60S maturation and release of several preribosome maturation factors; may be involved in degradation of aberrant mRNAs
YNL132W/KRE33	Essential protein, required for biogenesis of the small ribosomal subunit; heterozygous mutant shows haploinsufficiency in K1 killer toxin resistance
YMR078C/CTF18	Subunit of a complex with Ctf8p that shares some subunits with Replication Factor C and is required for sister chromatid cohesion; may have overlapping functions with Rad24p in the DNA damage replication checkpoint
YLR422W/YLR422W	Protein of unknown function with similarity to human DOCK proteins (guanine nucleotide exchange factors); interacts with Ino4p; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm, YLR422W is not an essential protein

YMR125W/STO1	Large subunit of the nuclear mRNA cap-binding protein complex, interacts with Npl3p to carry nuclear poly(A)+ mRNA to cytoplasm; also involved in nuclear mRNA degradation and telomere maintenance; orthologous to mammalian CBP80
YOR371C/GPB1	Multistep regulator of cAMP-PKA signaling; inhibits PKA downstream of Gpa2p and Cyr1p, thereby increasing cAMP dependency; promotes ubiquitin-dependent proteolysis of Ira2p; regulated by G-alpha protein Gpa2p; homolog of Gpb2p
YMR094W/CTF13	Subunit of the CBF3 complex, which binds to the CDE III element of centromeres, bending the DNA upon binding, and may be involved in sister chromatid cohesion during mitosis
YMR167W/MLH1	Protein required for mismatch repair in mitosis and meiosis as well as crossing over during meiosis; forms a complex with Pms1p and Msh2p-Msh3p during mismatch repair; human homolog is associated with hereditary non-polyposis colon cancer
YDR103W/STE5	Pheromone-response scaffold protein that controls the mating decision; binds Ste11p, Ste7p, and Fus3p kinases, forming a MAPK cascade complex that interacts with the plasma membrane and Ste4p-Ste18p; allosteric activator of Fus3p
YDR318W/MCM21	Protein involved in minichromosome maintenance; component of the COMA complex (Ctf19p, Okp1p, Mcm21p, Ame1p) that bridges kinetochore subunits that are in contact with centromeric DNA and the subunits bound to microtubules
YAL026C/DRS2	Aminophospholipid translocase (flippase) that maintains membrane lipid asymmetry in post-Golgi secretory vesicles; contributes to clathrin-coated vesicle formation and endocytosis; mutations in human homolog ATP8B1 result in liver disease
YDR180W/SCC2	Subunit of cohesin loading factor (Scc2p-Scc4p), a complex required for loading of cohesin complexes onto chromosomes; involved in establishing sister chromatid cohesion during DSB repair via histone H2AX; evolutionarily-conserved adherin
YOR092W/ECM3	Non-essential protein of unknown function; involved in signal transduction and the genotoxic response; induced rapidly in response to treatment with 8-methoxypsoralen and UVA irradiation
YDR235W/PRP42	U1 snRNP protein involved in splicing, required for U1 snRNP biogenesis; contains multiple tetriatricopeptide repeats
YER151C/UBP3	Ubiquitin-specific protease that interacts with Bre5p to co-regulate anterograde and retrograde transport between the ER and Golgi; inhibitor of gene silencing; cleaves ubiquitin fusions but not polyubiquitin; also has mRNA binding activity
YMR275C/BUL1	Ubiquitin-binding component of the Rsp5p E3-ubiquitin ligase complex, functional homolog of Bul2p, disruption causes temperature-sensitive growth, overexpression causes missorting of amino acid permeases
YKL114C/APN1	Major apurinic/apyrimidinic endonuclease, 3'-repair diesterase involved in repair of DNA damage by oxidation and alkylating agents; also functions as a 3'-5' exonuclease to repair 7,8-dihydro-8-oxodeoxyguanosine
YOL081W/IRA2	GTPase-activating protein that negatively regulates RAS by converting it from the GTP- to the GDP-bound inactive form, required for reducing cAMP levels under nutrient limiting conditions, has similarity to Ira1p and human neurofibromin
YPR049C/ATG11	Adapter protein for pexophagy and the cytoplasm-to-vacuole targeting (Cvt) pathway; directs receptor-bound cargo to the phagophore assembly site (PAS) for packaging into vesicles; required for recruiting other proteins to the (PAS)
YGL095C/VPS45	Protein of the Sec1p/Munc-18 family, essential for vacuolar protein sorting; required for the function of Pep12p and the early endosome/late Golgi SNARE Tlg2p; essential for fusion of Golgi-derived vesicles with the prevacuolar compartment
YDR456W/NHX1	Na+/H+ and K+/H+ exchanger, required for intracellular sequestration of Na+ and K+; located in the vacuole and late endosome compartments; required for osmotolerance to acute hypertonic shock and for vacuolar fusion
YKL197C/PEX1	AAA-peroxin that heterodimerizes with AAA-peroxin Pex6p and participates in the recycling of peroxisomal signal receptor Pex5p from the peroxisomal membrane to the cystosol; induced by oleic acid and upregulated during anaerobiosis
YIL068C/SEC6	Essential 88kDa subunit of the exocyst complex, which mediates polarized targeting of secretory vesicles to active sites of exocytosis; dimeric form of Sec6p interacts with Sec9p in vitro and inhibits t-SNARE assembly

YOR326W/MYO2	One of two type V myosin motors (along with MYO4) involved in actin-based transport of cargos; required for the polarized delivery of secretory vesicles, the vacuole, late Golgi elements, peroxisomes, and the mitotic spindle
YNR045W/PET494	Mitochondrial translational activator specific for the COX3 mRNA, acts together with Pet54p and Pet122p; located in the mitochondrial inner membrane
YJR107W/YJR107W	Putative protein of unknown function; has sequence or structural similarity to lipases
YPL268W/PLC1	Phospholipase C, hydrolyzes phosphatidylinositol 4,5-biphosphate (PIP2) to generate the signaling molecules inositol 1,4,5-triphosphate (IP3) and 1,2-diacylglycerol (DAG); involved in regulating many cellular processes
YJL062W/LAS21	Integral plasma membrane protein involved in the synthesis of the glycosylphosphatidylinositol (GPI) core structure; mutations affect cell wall integrity
YCR042C/TAF2	TFIID subunit (150 kDa), involved in RNA polymerase II transcription initiation

Supplementary Table 7. Annotations of genes from thermotolerance RH-seq. Shown are hits from thermotolerance mapping by RH-seq (Table S8) that met quality control thresholds and at which disruption of the *S. cerevisiae* allele compromised thermotolerance to a greater extent than did disruption of the *S. paradoxus* allele in the interspecific hybrid.

Supplementary Table 8. Effects of allelic variation in thermotolerance RH-seq. Each row reports the results of reciprocal hemizygote tests of thermotolerance of hemizygote transposon mutants at the indicated gene in the *S. cerevisiae* x *S. paradoxus* diploid hybrid at 37°C. Columns B-G report analyses upon the aggregation at the gene level of transposon mutant abundances (Table S5) from all biological replicate experiments, filtered for quality control (see Methods). Columns B-D report results of a two-tailed Mann-Whitney statistical test for a difference in the abundance after growth at 37°C, relative to the abundance after growth at 28°C, of hemizygotes harboring transposon insertions in the two species parents' homologs. The Benjamini-Hochberg method was used to correct for multiple testing. Column E reports the $\log_2$(abundance at 37°C/abundance at 28°C) of the average insert in the *S. cerevisiae* allele. Column F reports the analogous quantity among inserts in the *S. paradoxus* allele of the gene. Column G reports the allele-specific effect size, calculated as the difference between the measures of Columns E and F.

Supplementary Table 9. **Effects of allelic variation in RH-seq at 36°C.** Data are as in Table S8, except that the high temperature growth condition was 36°C.

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

Supplementary Table 10. Whole-gene tests for evidence of non-neutral protein evolution at thermotolerance loci. Each row reports results from the McDonald-Kreitman test on sequences from strains of European populations of *S. cerevisiae* and *S. paradoxus* of the indicated thermotolerance locus. D_s, number of sites of synonymous nucleotide divergence between species; number of sites of D_n,

nonsynonymous nucleotide divergence between species; P_s, number of sites of synonymous nucleotide polymorphisms within species; P_n, number of sites of nonsynonymous nucleotide polymorphisms within species. NI, neutrality index. The sixth column reports the *p*-value from a Fisher's exact test on D_s, D_n, P_s, and P_n, and the seventh column reports the adjusted *p*-value after applying the Benjamini-Hochberg correction for multiple hypothesis testing. All loci exhibited NI > 1, corresponding to a dearth of divergent amino acid changes relative to synonymous changes and polymorphisms.