





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

Melanie B. Abrams ^{1,†}, Julie N. Chuong ^{1,2,†}, Faisal AlZaben,^{1,†} Claire A. Dubin ^{1,†}, Jeffrey M. Skerker,^{3,*,†} and Rachel B. Brem ^{1,4,*}

¹Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA,

²PhD Program in Biology, New York University, New York, NY 10003, USA,

³Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, and

⁴Buck Institute for Research on Aging, Novato, CA 94945, USA

*Corresponding author: Department of Plant and Microbial Biology, University of California, Berkeley, 312F Innovative Genomics Institute Building, 2151 Berkeley Way, Berkeley, CA 94730, USA. Email: rbrem@berkeley.edu (R.B.); Zymergen, Inc., 5980 Horton St #105, Emeryville, CA 94608, USA. Email: skerker1@gmail.com (J.S.)

[†]Present address: Zymergen, Inc., 5980 Horton St #105, Emeryville, CA 94608, USA.

[‡]These authors contributed equally to this work.

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

Keywords: evolution; genetics; *Saccharomyces*; thermotolerance; adaptation

Introduction

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

The budding yeast *Saccharomyces cerevisiae* grows better at high temperature than any other species in its clade (Sweeney et al. 2004; Gonçalves et al. 2011; Salvadó et al. 2011; Hittinger 2013; Weiss et al. 2018), in keeping with its likely ecological origin in hot, East Asian locales (Peter et al. 2018). This derived and putatively adaptive trait serves as a model for the genetic study of

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

Received: July 24, 2021. Accepted: November 04, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

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

Materials and methods

Construction of a randomly barcoded piggyBac transposase pool

For barcoded RH-seq, we constructed a pool of plasmids, each harboring the piggyBac transposase and a randomly barcoded copy of the piggyBac transposon, via golden-gate cloning of random 20bp barcodes flanked by universal priming sites into a plasmid backbone containing the piggyBac machinery, modified from pJR487 (Weiss *et al.* 2018) as follows (Supplementary 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 (Supplementary Table S1); see Supplementary Note and Figure S2 for a description of this choice. The stuffer contained two BbsI cut sites with custom type IIS overhang sequences from Lee *et al.* (2015), and a NotI cut site in between the two BbsI cut sites. All cloning steps were carried out by GenScript, Inc.

Preparation of barcode oligonucleotides

To make barcodes, we acquired an oligonucleotide pool from IDT that contained random 20bp 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 the respective universal priming sequence (Supplementary Table S2) (Coradetti

et al. 2018). We set up 50 μ l amplification PCR reactions with 1 μ l of random 20bp barcodes as template, from a 2.5- μ M stock, and 0.25 μ l of each of the forward and reverse primers from a 100- μ M stock. Amplification used Phusion High Fidelity polymerase (NEB) and the following cycling protocol: 98°C for 30s (98°C for 10s, 58°C for 30s, and 72°C for 60s) \times 6, 72°C for 5 min. 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 μ g of pJC31 and 128ng of donor barcodes), 5 μ l of 10 \times T4 Ligase Buffer (ThermoFisher), 2.5 μ l of T4 Ligase (ThermoFisher), 2.5 μ l FastDigest Bpil (ThermoFisher), and sterile water up to 50 μ l. The cycling program was: 37°C for 5 min (37°C for 2 min, 16°C for 5 min) \times 25, 65°C for 10 min. Then a mixture containing 5 μ l 10 \times FastDigest Buffer (ThermoFisher), 3.13 μ l BSA 2 mg/ml (NEB), 12.5 μ l FastDigest NotI (ThermoFisher), and 12.5 μ l FastDigest Bpil (ThermoFisher) was spiked into the reaction and incubated at 37°C for 16 h to digest unbarcoded backbone vectors. Ten of these reactions were combined, purified, and eluted in H₂O (Zymo DNA Clean & Concentrator). To spot-check this cloning, 5 μ l of this product was transformed into 25 μ l of *E. coli* 10-beta electrocompetent cells (NEB). Sanger sequencing across the barcode regions of 20 individually minipreped *E. coli* colonies showed 95% barcoding efficiency.

The second reaction contained 2:1 molar ratio of vector to donor barcodes (4 μ g of pJC31 and 128ng of donor barcodes), 5 μ l of 10 \times T4 Buffer (ThermoFisher), 2.5 μ l T4 Ligase (ThermoFisher), 2.5 μ l Bpil (ThermoFisher), and sterile water up to 50 μ l. The cycling program was: 37°C for 5 min (37°C for 2 min, 16°C for 5 min) \times 25, 65°C for 10 min. Then a mixture containing 2.5 μ l 10 \times FastDigest Buffer (ThermoFisher), 2.5 μ l G Buffer, (ThermoFisher), 3.13 μ l BSA 2 mg/ml (NEB), 12.5 μ l FastDigest NotI (ThermoFisher), and 12.5 μ l Bpil (ThermoFisher) was spiked in the reaction and incubated at 37°C for 16 h to digest remaining unbarcoded backbone vectors. Six of these reactions were combined, purified, and eluted in H₂O (Zymo DNA Clean & Concentrator). Then every 5 μ l of cleaned eluted product was redigested with 5 μ l of NotI-HF (NEB), 5 μ l 10 \times CutSmart buffer (NEB), and 35 μ l H₂O at 37°C for 16 h then 80°C for 20 min. The reactions were purified again (Zymo DNA Clean & Concentrator) and pooled. Spot checks of this cloning reaction proceeded as above, and Sanger sequencing across the barcode regions of 20 individually minipreped *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 μ l of 10-beta electrocompetent cells (NEB) and 5 μ l 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 2.0kV, 200 Ω , 25 μ F. After electroporation, each culture was recovered in provided outgrowth media (NEB) by shaking at 37°C at 250 rpm for 1.5 h. After recovery, all independent 21 electroporation reactions were combined.

The combined recovered transformation *E. coli* culture was used to inoculate two 11 fresh LB cultures containing

carbenicillin at 100 µg/ml to select for *E. coli* cells containing bar-coded piggyBac plasmids. Each culture was incubated for 15.5 h 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 h at 37°C, 250 rpm or until it reached an OD₆₀₀ of 2.1. This culture was gigaprepmed on using a column kit (Invitrogen) to generate 5 mg of plasmid. We used this as input into a PCR with primers (Supplementary 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 AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACAGGACGCTCTTCCGATCT(N1–4)xxxxxxGTCGACCTGCAGCTACG, where the N1–4 represent variable amounts of random bases from 1 to 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 prepmed P58 plasmid template each. Amplification used Q5 High Fidelity Polymerase (NEB) and a cycling program 98°C for 4 min (98°C for 30 s, 55°C for 30 s, 72°C for 30 s) × 25, 72°C for 5 min. 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 microliters 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 Supplementary 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.

To prepare plasmid DNA for mutagenesis, one frozen aliquot of P58 was inoculated into 1.25 l of LB containing carbenicillin 100 µg/ml and grown for 16 h at 37°C, 250 rpm or until it reached an OD₆₀₀/ml of 2.1. This culture was gigaprepmed on using a column kit (Invitrogen) to generate 5 mg of plasmid.

Next, we transformed yeast in several, smaller subpools that 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 two days. The first day, we generated one subpool in a single 50 ml culture and one subpool in five 50 ml cultures at OD₆₀₀/ml ~0.9 (~45 OD₆₀₀ units of cells each). The second day, we generated two subpools in five 50 ml cultures each at OD₆₀₀/ml ~0.9 (~45 OD₆₀₀ 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. Twenty-four hours later, the OD₆₀₀/ml of the

overnight culture was 3.86. It was back-diluted to an OD₆₀₀/ml of 0.1 in 50 ml of YPD in a 250-ml Erlenmeyer flask and grown with shaking at 28°C, 200 rpm for 5.5 h. After 5.5 h, it had reached OD₆₀₀/ml ~0.9 and cells were at mid-log phase. This 50 ml culture was gently pelleted at 1000×g for 3 min. The pellet was washed with 25 ml sterile water and then 5 ml of 0.1 M lithium acetate (Sigma) mixed with 1× 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 1.2× Tris–EDTA buffer (12 mM Tris–HCl and 1.2 mM EDTA). The tube was rested for 10 min at room temperature, then heat-shocked in a water bath at 37°C for 26 min. The tube was gently spun at 1000×g for 3 min after which supernatant was removed. We transferred the cells to a flask and added YPD to attain an OD₆₀₀/ml 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 OD₆₀₀/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 OD₆₀₀/ml of 0.2 in 50 ml each. We cultured each flask with 200 rpm shaking at 28°C overnight until each reached an OD₆₀₀/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 OD₆₀₀/ml. Four milliliters of this resuspension were plated (1 ml per 24.1 cm × 24.1 cm plate) onto plates containing complete synthetic media with 5-fluoroorotic 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-fluoroorotic acid (Zymo Research)]. After incubation at 28°C to enable colony growth, colonies were scraped off all four plates and combined into water at the volume required to attain 44 OD₆₀₀/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 OD₆₀₀/ml of the overnight culture was OD₆₀₀/ml 3.89. The overnight culture was back-diluted to OD₆₀₀/ml 0.1 in 250 ml of YPD and incubated for 5.5 h at 28°C, 200 rpm. After 5.5 h, the OD₆₀₀/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 1 l flask and added fresh YPD to attain OD₆₀₀/ml 0.4 in ~750 ml YPD. The transformed culture was recovered by shaking at 28°C, 200 rpm, for 2 h. G418 (300 mg/µl) was added to select for the transposed cells. The culture continued shaking for 48 h or until the OD₆₀₀/ml reached 2.1. This culture was then back-diluted to create a new culture at OD₆₀₀/ml 0.2 in 500 ml of YPD with 300 mg/µl G418 shaking for 24 h at 28°C, 200 rpm until it reached OD₆₀₀/ml ~3.4. The curing, scraping, and freezing steps were the same as above.

To combine the four subpools to yield the final 160× hemizygote pool (final identifier P75), three 1 ml aliquots of each subpool were thawed on ice for 1 h. They were transferred to each of four 1 L flasks with 500 ml YPD to OD₆₀₀/ml 0.2, cultured at 28°C, 200 rpm for 17 h after which the OD₆₀₀/ml was 3.5–4. These cultures were gently pelleted, combined, and resuspended either in

YPD with 15% glycerol or YPD with 7% DMSO, to a final OD₆₀₀/ml of 44; 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 44 OD₆₀₀/ml aliquot of each subpool or final pool was thawed on ice, and its genomic DNA (gDNA) was harvested with the ZR Fungal/Bacterial DNA MiniPrep Kit (Zymo Research). gDNA was resuspended in DNA elution buffer (Zymo Research) prewarmed to 65°C , and its concentration was quantified using a Qubit 4.0 fluorometer. Illumina transposon sequencing (Tn-seq) library construction was as described. 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 Supplementary Table S2 for the transposon-specific primer (“forward primer”), where Ns represent random nucleotides, and the indexed adapter-specific primer (“reverse primer”). Amplification used Jumpstart polymerase (Sigma) and the following cycling protocol: 94°C for 2 min (94°C for 30 s, 65°C for 20 s, 72°C for 30 s) \times 25, 72°C for 10 min. Sequencing of paired-end reads of 150 bp was done over two lanes on a HiSeq4000 at Novogene Corporation (Sacramento, CA, USA) and one lane on a NovaSeq SP at the Genomics Sequencing Laboratory at UC Berkeley (Berkeley, CA, USA). Reads sequenced per library are reported in Supplementary Table S4.

Tn-seq data analysis

Tn-seq data of the hemizygote pool was analyzed, to infer transposon insertions on the basis of barcodes detected in reads as junctions with genomic sequence, essentially as described (Coradetti *et al.* 2018) (https://github.com/stcoradetti/RBseq/tree/master/Old_Versions/1.1.4; last accessed October 10, 2021), with the following modifications. For each barcode, instead of scanning positions for the end of the insertion from a sequence specified by a model file, we searched for the final 22 base pairs of the right arm of the piggyBac transposon allowing for two mismatches. For annotation, we converted the annotation file from <https://github.com/weiss19/rh-seq> (last accessed June 5, 2021) for the *S. cerevisiae* D1373 \times *S. paradoxus* Z1 hybrid to a compliant GFF3 file using Another GFF Analysis Toolkit (AGAT) - Version: v0.4.0 (<https://github.com/NBISweden/AGAT>). Then, we used a custom Jupyter notebook to annotate the file generated by the RBseq mapping software.

Quality control for Tn-seq, to eliminate barcodes whose junction genomic sequence mapped to multiple insertion locations in the hybrid genome, and to minimize the proportion of sequencing errors included in final tallies, was as described (Coradetti *et al.* 2018). Briefly, we eliminated from further consideration any case where a barcode observed in Tn-seq sequencing data differed from another, much more abundant, barcode by a single base (a total of 2,024,812 off-by-one barcodes in 2,888,129 reads). We also filtered out off-by-two barcodes (280,949 barcodes in total). Separately, we eliminated barcodes that were detected in sequencing data as a junction with more than one genomic context, suggesting the respective transposon had inserted into

multiple locations in one or many clones (98,669 barcodes where this inference was based on multiple strong mapping matches, and an additional 46,583 barcodes where this inference was ambiguous, with one strong mapping match with reads outnumbered by those assigned to weaker mapping matches). The final filtered barcode set comprised 548,129 uniquely barcoded and mapped inferred transposon insertions in the P75 hemizygote pool, at an average read depth of 308.6 reads, and a median read depth of 47 reads; 166,834 of these insertions were mapped as genic. The annotation script, GFF3 file, and modified mapping script are available at https://github.com/melanieabrams-pub/RH-seq_with_barcoding.

Competition cultures

For thermotolerance competition at 37°C (Supplementary Table S5), 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 6 h at 28°C , 200 rpm. This pre-culture (T_0 , at OD₆₀₀/ml of 1.22) was back-diluted into 12 10 ml competition cultures at 200 rpm at each of 28°C and 37°C , with a starting OD₆₀₀/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 h after the T_0 timepoint, and dilutions for the 37°C competition cultures were performed after 8.5, 18.5, 25.5, and 32.5 h after the T_0 timepoint. The entire cell culture was harvested from each of these biological replicate tubes for sequencing as biological replicates.

For thermotolerance competition at 36°C (Supplementary Table S6), competition cultures were grown as above with the following differences. The high temperature was 36°C , instead of 37°C . The preculture (T_0 , at OD₆₀₀/ml of 0.693 after 5.5 h at 28°C , 200 rpm) was back-diluted to a starting OD₆₀₀/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 h after the T_0 timepoint. Eleven instead of 12 replicates were carried out at 28°C .

Barcode quantification from competition cultures

Bar-seq library preparation

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

Bar-seq data analysis

Bar-seq mapping and quantification were as described (Coradetti *et al.* 2018) (https://github.com/stcoradetti/RBseq/tree/master/Old_Versions/1.1.4; last accessed October 10, 2021), wherein only barcodes that passed quality control in Tn-seq (see *Tn-seq data*

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

Competitions for the dissection of growth at 36°C relative to 28°C (Supplementary Table S6) used the same procedures as above, mapping a total of 230,469 barcodes, 68,523 of which corresponded to inserts in genes and were analyzed as input to the reciprocal hemizyosity testing pipeline described below. In a given replicate competition culture, we detected a median 5×10^4 barcodes.

Reciprocal hemizyosity testing

The tabulated counts of abundance from Bar-seq for each barcode in each replicate were used as input into reciprocal hemizyosity tests essentially as in [Abrams et al. \(2021\)](#) with slight changes as follows. We had in hand each barcode which had been sequenced as a junction with a unique genomic location in the Tn-seq step and had passed quality control there (see *Tn-seq data analysis*), and which was now detected in competition cultures. We interpreted each such barcode as reporting a hemizygote clone bearing a transposon insertion at the respective position of the respective species' allele (*S. cerevisiae* or *S. paradoxus*), with the other species' allele retained as wildtype at that locus. In what follows, we refer to each such barcode as reporting an inferred hemizygote clone, with respect to its growth behavior in competition cultures. As in [Abrams et al. \(2021\)](#), for a given biological replicate we normalized the abundances attributed to each inferred hemizygote genotype to the total number of sequencing reads in the respective sample, and we eliminated from further analysis insertions which had been annotated as intergenic, or as corresponding to the plasmid used to generate this pool. For reciprocal hemizyosity tests, we excluded from consideration any gene with fewer than three inferred hemizygote genotypes per allele. Of the retained genes, for each inferred hemizygote genotype, we tabulated the quantity $a_{\text{experimental},i}$, the sequencing-based abundance measured after the competition culture in biological replicate i of growth at the experimental temperature (36°C or 37°C), and, separately, we calculated $a_{\text{control},i}$, the analogous quantity from growth at the control temperature (28°C), for $i = [1,12]$. We then took the mean of the latter and used it to tabulate the temperature effect on the inferred hemizygote genotype in replicate i , $t_i = \log_2(a_{\text{experimental},i}/a_{\text{control}, \text{mean}})$. As in [Abrams et al. \(2021\)](#), we eliminated an inferred hemizygote genotype if the coefficient of variation of this quantity exceeded 2.0, or there were fewer than 1.1 normalized reads. With the data for the remaining inferred hemizygote genotypes (Supplementary Tables S5 and S6), for a given gene, we compiled the vector of the t measurements across all replicates and all inferred hemizygote genotypes with each species' allele of the

hybrid disrupted in turn, and discarded genes where the coefficient of variation of the t measurements across hemizygote inserts for one or both alleles exceeded 10. For the remainder, we used the Mann–Whitney test to compare these two vectors, with Benjamini–Hochberg correction for multiple testing (Supplementary Tables S7 and S8). For a given gene, we calculated the effect size as the difference between two values: the $\log_2(\text{abundance at the experimental temperature/abundance at } 28^\circ\text{C})$ of the average inferred hemizygote genotype representing a transposon insertion in the *S. cerevisiae* allele, and the analogous quantity among inferred hemizygote genotypes representing insertions in the *S. paradoxus* allele of the gene. Scripts for this modified RH-seq analysis pipeline are available at https://github.com/melanieabrams-pub/RH-seq_with_barcoding. We earmarked top candidate genes for factors contributing to the thermotolerance of *S. cerevisiae* as those with corrected Mann–Whitney $P < 0.05$ in the reciprocal hemizyosity test, and an effect size < -0.5 , i.e., disrupting the *S. cerevisiae* allele was associated with a strong defect in thermotolerance relative to disruption of the *S. paradoxus* allele; we refer to this gene set as our top barcoded RH-seq hit gene list.

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

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

To test the encoded proteins of top barcoded RH-seq hit genes for enrichment of physical interactions with each other, we used curated known interactions from BioGRID ([Oughtred et al. 2021](#)) as housed in the *Saccharomyces* Gene Database, downloaded February 19, 2021. We tabulated the number of physical interactions between the proteins encoded by RH-seq hit genes, and we divided that by the total number of interactions involving one RH-seq hit gene and any other gene in the genome; call this ratio r_{true} . Then, we drew a random sample of genes from the genome, as described above for GO term resampling. We tabulated, in this random gene set, the number of physical interactions between genes in that sample, and we divided that by the total number of interactions involving one gene in the random sample and any other gene in the genome, to yield r_{resample} . We repeated this procedure 10,000 times, and we used the proportion of resampled groups where r_{resample} was greater than or equal to r_{true} as a one-sided P -value assessing the significance of enrichment of interactions between our genes of interest.

GO analyses of top hit genes from barcoded RH-seq

To test top barcoded RH-seq hit genes for enrichment for overrepresentation of a particular GO term, we mapped each gene to its GO groups based on data from geneontology.com ([Ashburner et al. 2000](#)), last accessed May 13, 2021. We filtered out GO terms with fewer than five or with more than 200 gene members. We also filtered out GO terms with identical membership in the genome. We took the subset of the remaining GO terms with at least one member among our top barcoded RH-seq hit genes. For each such term, we randomly sampled genes from the genome, ensuring the same proportion of essential genes as in our set of

top barcoded RH-seq hit genes based on the essentiality annotations of Winzeler (1999). We tabulated whether our random sample had greater or fewer genes with the term of interest than our candidate set. We repeated this procedure 10,000 times and used the proportion of these resampled groups that had more genes in the term as the initial P -value assessing the significance of the enrichment of that term. Then, we applied Benjamini–Hochberg correction for multiple hypothesis testing across all terms to generate final, adjusted P -values across the experiment.

To test Biological Process ontologies for enrichment for large magnitudes of the effect of allelic variation on thermotolerance, we used the latter as tabulated in *Reciprocal hemizyosity testing*. We filtered GO terms as above, and then excluded all genes absent in our barcoded RH-seq analysis. For each retained term in turn, we first tabulated the median absolute value of the effect size of the gene members for which we had data, e_{true} . Then, we tabulated the analogous quantity for a random sample of the same number of genes from the genome, e_{resample} , ensuring the same proportion of essential genes as above. We repeated this procedure 100 times, and used the proportion of the resampled groups for which e_{resample} was greater than or equal to e_{true} as an initial P -value assessing the enrichment of large effects of allelic variation in the genes the term. For all GO terms with an initial P -value < 0.1 , we repeated this procedure 10,000 times to calculate a more precise P -value. Then, we applied the Benjamini–Hochberg correction for multiple hypothesis testing to generate final, adjusted P -values across the experiment.

Molecular evolution analysis of RH-seq hit genes

Branch length PAML analysis with codeML was performed as in Dubin et al. (2020). Hits were manually inspected for the quality of the alignment, and one, YAL026C, was discarded for poor alignment quality leading to an artifactually high branch length. We used the inferred branch lengths as input into a resampling test as in *GO analyses of top hit genes from barcoded RH-seq*, and we performed a one-sided significance test for long branch lengths along the *S. cerevisiae* lineage. Branch-site PAML analysis with codeML was performed as in Abrams et al. (2021). Jalview Version 2 was used to visualize the percentage of identity of amino acid sequence alignments (Waterhouse et al. 2009). 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.

Results

Dissecting thermotolerance divergence between species by barcoded transposon mutagenesis

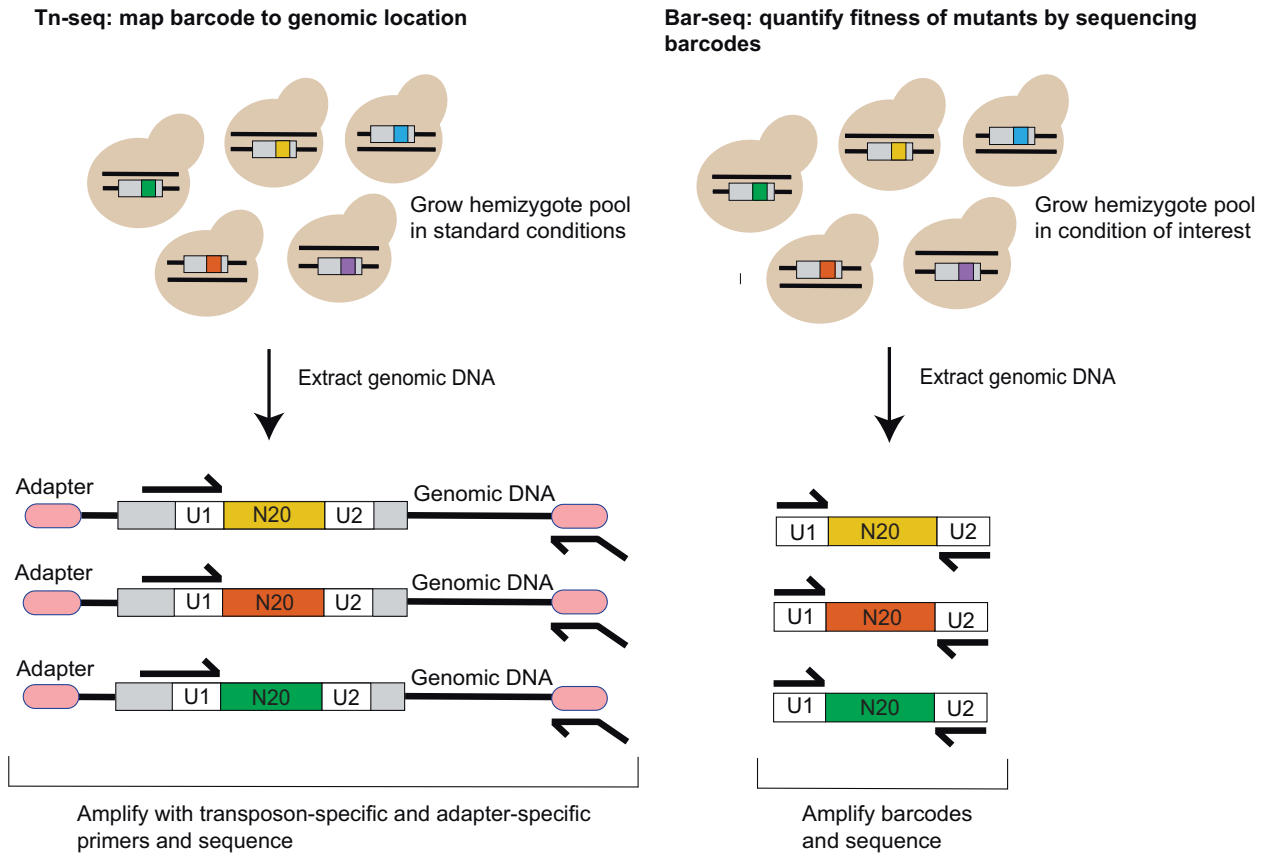
With the goals of boosting RH-seq throughput and power, and achieving new insights into the genetics and evolution of yeast thermotolerance, we set out to generate an RH-seq reagent for yeast incorporating barcoded transposons (Wetmore et al. 2015). For this purpose, we first generated a pool of plasmids, each encoding a barcoded copy of the piggyBac transposon and its transposase (Supplementary Figure S1, A–C). To use these in RH-seq, we revisited our previously characterized model system: a comparison between DVBP1373, a thermotolerant Dutch soil strain of *S. cerevisiae*, and Z1, an *S. paradoxus* isolate from the United Kingdom (Weiss et al. 2018; Abrams et al. 2021; AlZaben et al. 2021). The F1 hybrid formed from the mating of these strains exhibits a thermotolerance phenotype intermediate between those of the two species parents, and thus is well suited to

mapping of allelic effects on the trait (Weiss et al. 2018). We transformed this F1 hybrid with barcoded plasmids, yielding a pool of hemizygote mutants, which we expanded and then banked (Supplementary Figure S1D). Next, to catalog the genomic locations of transposon insertions, we used the DNA from a culture of the pool in standard conditions as input into a first round of sequencing library construction, whose primers recognized a common site on the transposon and a common DNA adapter ligated to DNA fragment ends (“Tn-seq”; Figure 1A). Sequencing and data analysis, with quality controls to eliminate barcodes that could not be uniquely associated with a single transposon insertion location (see *Materials and methods*), yielded a catalog of 548,129 barcoded hemizygotes in the pool whose genomic insertion locations were tabulated. At this point, we could harness the pool for highly replicated screens, each of which could quantify hemizygote abundance in a condition of interest via relatively cheap and straightforward barcode sequencing (“Bar-seq”; Figure 1B).

Thus, with our barcoded hemizygote pool, we implemented an RH-seq screen to search for genes at which *S. cerevisiae* and *S. paradoxus* alleles drove differences in strain abundance at high temperature. For this, we subjected the pool to growth assays with 12 biological replicate cultures at 37°C, alongside controls at 28°C. We used DNA from each culture as input into barcode sequencing (Figure 1B). The resulting data revealed a total of 301,349 cases where a barcode, representing a hemizygote clone with a transposon insertion catalogued by Tn-seq (Figure 1A), was detectable in our growth assays. Transposon insertion positions corresponding to these informative barcodes were evenly split between *S. cerevisiae* and *S. paradoxus* alleles of genes throughout the F1 hybrid genome (Supplementary Figure S3). We took the normalized count of a given barcode in a sequencing data set as a report of the fitness of the respective hemizygote, i.e., its relative abundance after growth in the pool in the respective condition. We then used the complete set of such counts as the input into reciprocal hemizyosity tests to compare, for a given gene, the temperature-dependent abundance of strains harboring a disruption in the *S. cerevisiae* allele, relative to that of strains with the *S. paradoxus* allele disrupted. A pipeline for these tests, including filters for coverage and reproducibility and multiple testing correction (see *Materials and methods*), revealed 83 genes at a 5% false discovery rate (Figure 2; Supplementary Table S7). This contrasted with the much smaller set of eight genes at which species’ alleles drove differences in high-temperature growth, identified in our original nonbarcoded RH-seq approach (Weiss et al. 2018), which had involved only three biological replicates. The 10-fold increase in the number of significant hits in our barcoded RH-seq screen reflects the statistical power afforded by our highly replicated method to detect even quite small effects.

In our barcoded RH-seq screen hits, as a positive control we first examined the set of genes known to contribute to thermotolerance divergence from our earlier study (AFG2, APC1, CEP3, DYN1, ESP1, MYO1, SCC2, and DYN1) (Weiss et al. 2018). Several did not meet the experiment-wide statistical thresholds of our barcoded RH-seq pipeline (Supplementary Figure S4A), suggesting an appreciable false negative rate of the latter overall. However, manual inspection made clear that hemizyosity effects at all gold-standard thermotolerance loci were borne out: in each case, in barcoded RH-seq data, strains with disruptions in the *S. cerevisiae* allele, and a wildtype copy of the *S. paradoxus* allele, had worse thermotolerance than did strains with only the *S. cerevisiae* allele intact (Supplementary Figure S4, A and B), as we had previously reported (Weiss et al. 2018). Furthermore, the list of gene hits from barcoded RH-seq also included HFA1

A



B

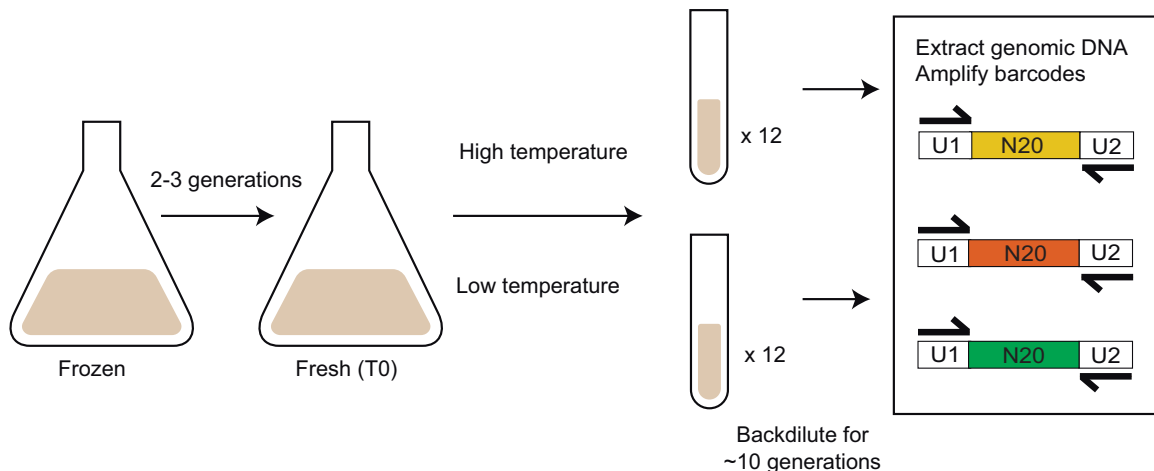


Figure 1 Barcoded RH-seq mapping of yeast thermotolerance loci. (A) Barcoded RH-seq sequencing analysis steps. Left, in a pool of *S. cerevisiae* × *S. paradoxus* hybrid hemizygotes, each harboring a transposon (gray rectangle) marked with a unique 20-mer barcode (multicolored) flanked by universal primer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

(Figure 2B; Supplementary Tables S7 and S9) which was reported and validated separately as a determinant of thermotolerance differences between yeast species (Li et al. 2019). On the strength

of these controls, we considered our deep sampling of thermotolerance loci to serve as a useful proof of concept for the barcoded RH-seq method.

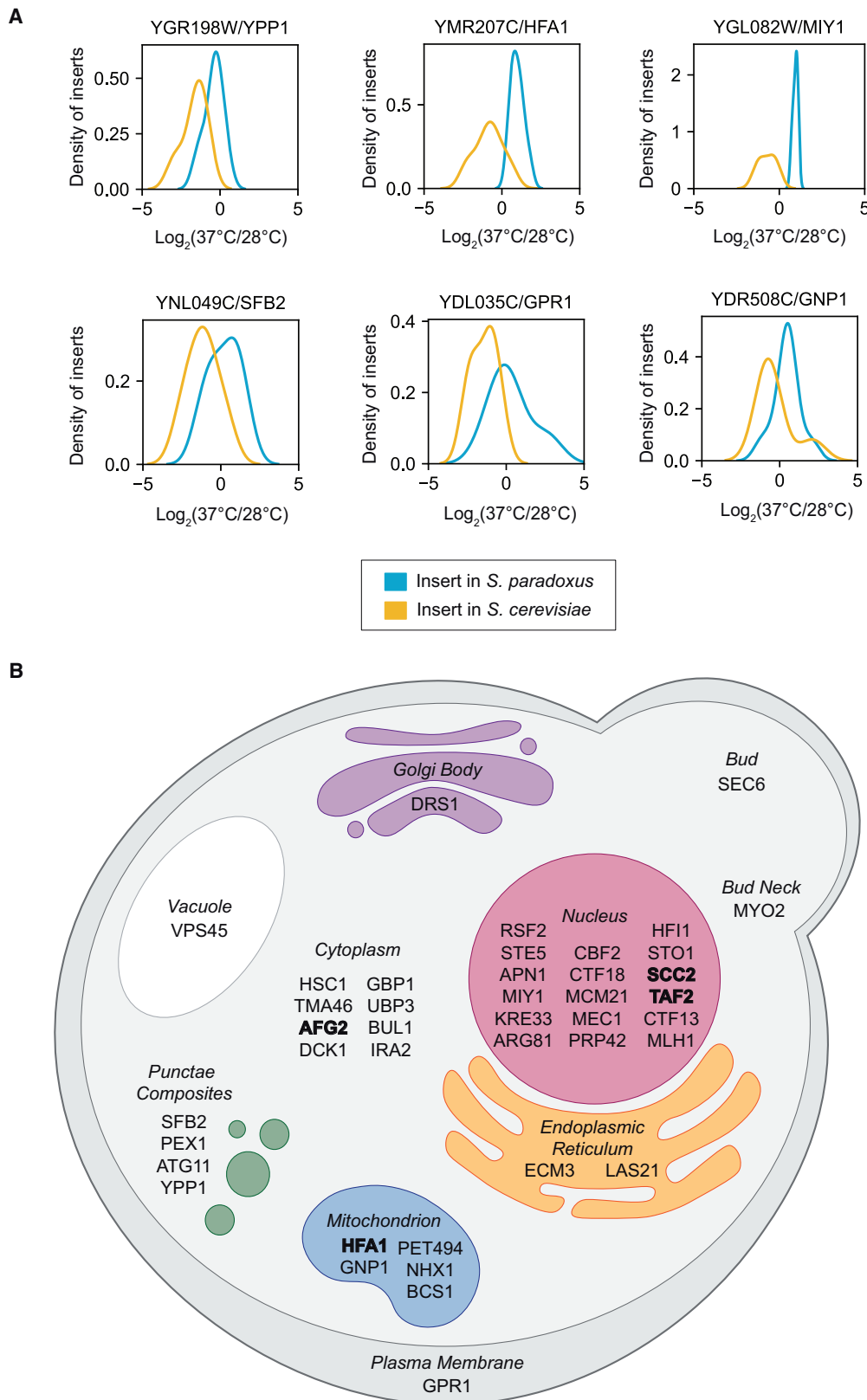


Figure 2 Hits from barcoded RH-seq mapping of yeast thermotolerance. (A) Each panel reports barcoded RH-seq results for a gene at which the *S. cerevisiae* allele was associated with better thermotolerance than the *S. paradoxus* allele, when uncovered in the hybrid background. In a given panel, the x-axis reports the \log_2 of abundance, measured by RH-seq after selection at 37°C, of a clone harboring a barcoded transposon insertion in the indicated species' allele in a given replicate, as a difference from the analogous quantity for that clone after selection at 28°C on average across replicates. The y-axis reports the proportion of observations of all clones bearing insertions in the indicated allele that exhibited the abundance ratio on the x, as a kernel density estimate. Shown are the top six genes from among all barcoded RH-seq hit loci in terms of allelic effect size; see Supplementary Table S7 for effect sizes of the complete set of hits. (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.

Functional-genomic analysis of thermotolerance genes

We next aimed to pursue deeper analyses of the novel gene hits from barcoded RH-seq in our yeast thermotolerance application. We considered that a focus on the strongest and most evolutionarily relevant sources of mapping signal would likely yield the most informative results. As such, in light of our interest in explaining the exceptional thermotolerance of purebred *S. cerevisiae*, we earmarked the 44 genes from our larger candidate set at which the *S. cerevisiae* allele boosted the trait most dramatically relative to that of *S. paradoxus* (Figure 2; Supplementary Table S9). In what follows, we refer to these genes as our top RH-seq hits, and we analyze them as our highest-confidence predictions for factors that nature would have used in evolving the *S. cerevisiae* phenotype.

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

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

Evolutionary analysis of thermotolerance genes

We anticipated that sequence analyses of the genes we had mapped to thermotolerance by barcoded RH-seq could shed light

on the evolutionary history of the trait. To explore this, we used a phylogenetic approach in *Saccharomyces sensu stricto*. We first inferred species-specific branch lengths in the phylogeny of each gene in turn, and focused on the lineage leading to *S. cerevisiae*. The distribution of branch lengths along this lineage among top thermotolerance gene hits was not detectably different from that of the genome as a whole, with the exception of two rapidly evolving thermotolerance genes, *TAF2* and *BUL1*, encoding a transcription initiation factor and ubiquitin ligase adapter, respectively (Supplementary Figure S5). Separately, we quantified protein evolutionary rates in top hits from barcoded RH-seq. A branch-site phylogenetic modeling approach (Yang 2007) detected striking evidence for positive selection along the *S. cerevisiae* lineage in the amino acid permease *GNP1*, the kinetochore DNA binding factor *CBF2*, and the sister chromatid cohesion factor *CTF18* (Figure 4). Interestingly, however, McDonald–Kreitman tests (McDonald and Kreitman 1991) on population-genomic data did not detect an overall excess of amino acid variation relative to synonymous changes, at these three genes or any other barcoded RH-seq hit locus (Supplementary Table S10). Thus, even at genes harboring individual codons with likely signatures of selection, we could not detect evidence for a scenario where *S. cerevisiae* stacked up a large number of unique amino acid changes, in the evolution of thermotolerance. Together, however, our analyses do highlight thermotolerance genes with marked signal for derived alleles in *S. cerevisiae* at single codons or in the overall DNA sequence—cases where species divergence is likely to be of particularly strong phenotypic and evolutionary importance.

Discussion

RH-seq power and the interpretation of mapped loci

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

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

By virtue of our focus on prothermotolerance alleles in *S. cerevisiae*, our work has left open the functional and evolutionary genomics of loci at which the allele from *S. cerevisiae* instead conferred worse thermotolerance than that of *S. paradoxus*, when each in turn was uncovered in the hybrid. Our barcoded RH-seq identified a number of such genes at high statistical significance.

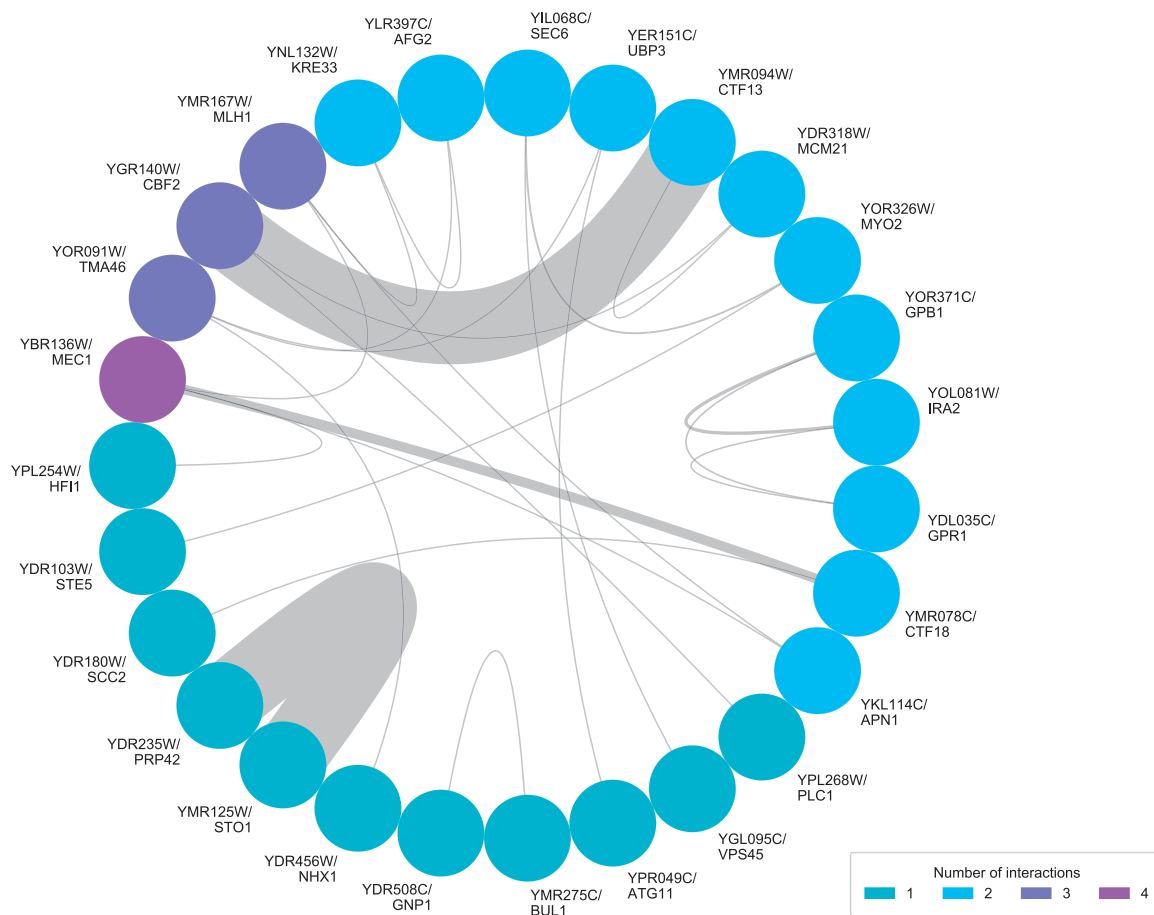


Figure 3 Interactions between thermotolerance loci. Each node represents a top hit gene from barcoded RH-seq mapping of thermotolerance. Each chord represents an inferred interaction, taking into account physical and genetic interactions as well as pathway membership, from the STRING database (Szklarczyk et al. 2021). Chords are weighted by the confidence of the inference of interactions; genes with higher numbers of interactions among the hits are represented by warmer colored nodes.

Table 1 Functional enrichment among thermotolerance loci

| GO term | $n_{\text{observed}}:n_{\text{expected}}$ | 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 |

Each row with numerical data reports a GO term enriched for RH-seq hit genes. n_{observed} , the number of genes from among top hits from thermotolerance RH-seq that were annotated with the term. n_{expected} , the number of genes annotated with the term in the same number of randomly chosen genes from the genome, as a median across samples. Adjusted P, resampling-based significance of the enrichment after Benjamini-Hochberg correction.

These loci may well reflect the accumulation of advantageous alleles in *S. paradoxus*, or deleterious alleles in *S. cerevisiae*, by drift, even as *S. cerevisiae* was under selection to improve the trait in evolutionary history. Analogously, in linkage mapping results, the effect of an allele in recombinant progeny from a cross often does not conform to that expected from the respective parent's phenotype (Burke and Arnold 2001; Brem and Kruglyak 2005). It is

also possible that some such allelic effects are the product of epistatic interactions between a locus of interest and the hybrid background, and would be phenotypically buffered (and thus evolutionarily irrelevant) in the purebred species. Molecular validation will be necessary to confirm the phenotypic impact of variation at our mapped loci, and its potential dependence on genetic background.

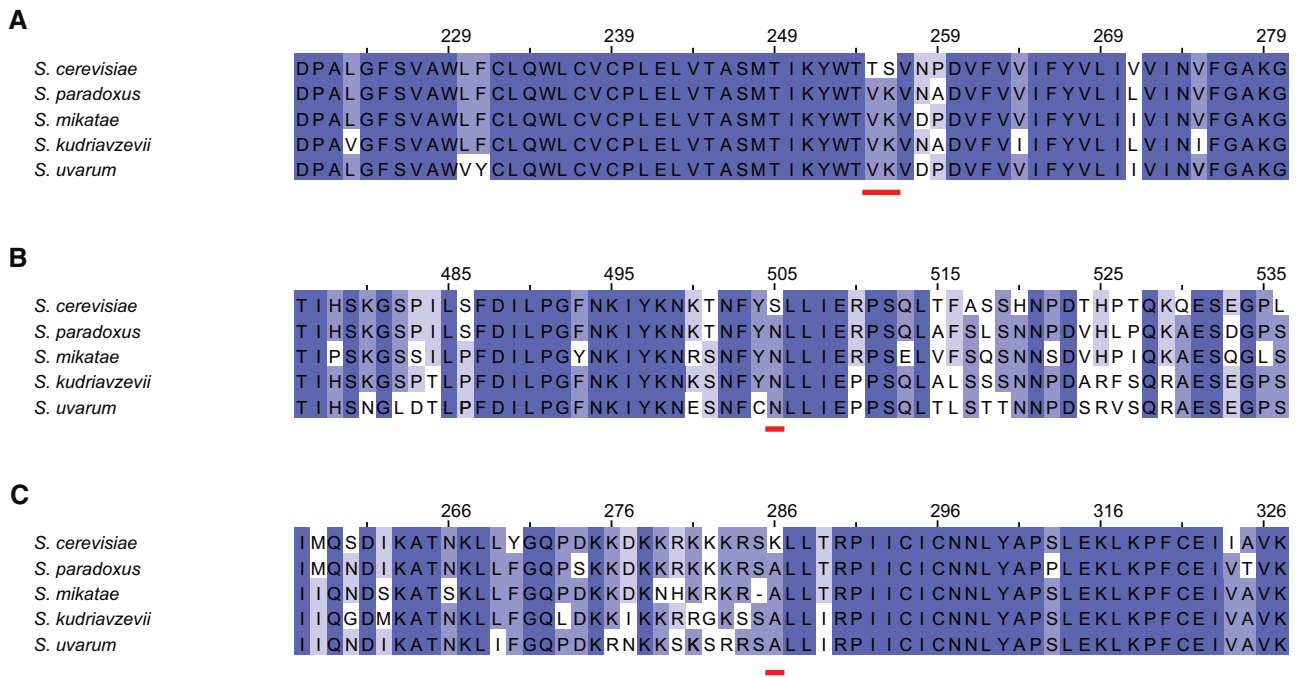


Figure 4 Codons under positive selection in thermotolerance loci. Each panel shows the amino acid sequence context, across type strains of *Saccharomyces sensu stricto* species, of codon(s) (red bar) inferred to be under positive selection along the *S. cerevisiae* lineage, in a hit gene from RH-seq thermotolerance mapping. Darker shading indicates a higher % identity. (A) YDR508C/GNP1, (B) YGR140W/CBF2, and (C) YMR078C/CTF18.

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

Cellular and molecular mechanisms of thermotolerance

Our top RH-seq hits revealed strong evidence for chromosome segregation and other mitosis functions as a linchpin of *S. cerevisiae* thermotolerance. As a complement to earlier characterization of six such genes (*APC1*, *ESP1*, *DYN1*, *MYO1*, *CEP3*, and *SCC2*) (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. Under one model of *S. cerevisiae* evolution, the particular niche to which this species specialized was one of avid fermentation, producing (and resisting) heat and ethanol that eliminated its microbial competitors (Goddard 2008; Salvadó et al. 2011). In such a scenario, the maximum benefit could well accrue to the organism if it were able to undergo rapid cell division under the challenging conditions of its own making. Consistent with this notion, another budding yeast, *Hanseniaspora*, which often dominates in early fermentation prior to takeover by *S. cerevisiae* (Fleet 2003), underwent evolutionary loss of much of the cell-cycle checkpoint

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

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

The latter idea emerged largely from a proteomic study which showed that thermotolerant organisms had higher thermostability of essential proteins of many functions, across the tree of life (Leuenberger et al. 2017). Were sequence changes that led to improved protein stability a linchpin of thermotolerance evolution in *S. cerevisiae*? Our data are consistent with a mechanistic role for properties of the protein sequences of many thermotolerance genes, in that variation in coding regions has come to the fore in our sequence tests here and those of an earlier small-scale analysis (Abrams et al. 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 non-coding 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 much of 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 mechanisms

in building the trait, given the apparent complexity of the problem. Biochemical studies will be necessary to nail down exactly how *S. cerevisiae* alleles advance thermotolerance in a given gene.

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

Data availability

Sequencing data are deposited in the Sequence Read Archive under the accession number PRJNA735401. Strains and plasmids are available upon request. Custom scripts for the barcoded RH-seq analysis are available at https://github.com/melanieabrams-pub/RH-seq_with_barcoding. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Supplementary material is available at figshare: <https://doi.org/10.25387/g3.16815100>.

Acknowledgments

The authors thank Adam Arkin for his generosity with computational resources; Morgan Price, Lori Huberman, and members of the John Dueber lab for discussions about barcoded transposon mutagenesis; and sequencing; and Abel Duarte for cloning advice.

Funding

This work was supported by National Science Foundation GRFP DGE 1752814 to M.A. and National Institutes of Health R01 GM120430 to R.B.B.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

- Abrams MB, Dubin CA, AlZaben F, Bravo J, Joubert PM, et al. 2021. Population and comparative genetics of thermotolerance divergence between yeast species. *G3 (Bethesda)*. 11:jkab139. doi:10.1093/g3journal/jkab139.
- AlZaben F, Chuong JN, Abrams MB, Brem RB. 2021. Joint effects of genes underlying a temperature specialization tradeoff in yeast. *PLoS Genet*. 17:e1009793. doi:10.1371/journal.pgen.1009793.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. 2000. Gene ontology: tool for the unification of biology. *Nat Genet*. 25:25–29. doi:10.1038/75556.
- Baker EP, Peris D, Moriarty RV, Li XC, Fay JC, et al. 2019. Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci Adv*. 5:eaav1869. doi:10.1126/sciadv.aav1869.
- Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 169:1177–1186. 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*. 102:1572–1577. 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, et al. 2018. Functional genomics of lipid metabolism in the oleaginous yeast *Rhodospiridium toruloides*. *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 (Bethesda)*. 10:2079–2085. doi:10.1534/g3.120.401304.
- Fleet GH. 2003. Yeast interactions and wine 16iggyb. *Int J Food Microbiol*. 86:11–22. doi:10.1016/S0168-1605(03)00245-9.
- Goddard MR. 2008. Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem engineering via fermentation. *Ecology*. 89:2077–2082. doi:10.1890/07-2060.1.
- Gonçalves P, Valério E, Correia C, de Almeida J, Sampaio JP. 2011. Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One*. 6:e20739. doi:10.1371/journal.pone.0020739.
- Hittinger CT. 2013. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet*. 29:309–317. doi:10.1016/j.tig.2013.01.002.
- Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, et al. 2003. Global analysis of protein localization in budding yeast. *Nature*. 425:686–691. 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*. 4:975–986. doi:10.1021/sb500366v.
- Leuenberger P, Ganschka S, Kahraman A, Cappelletti V, Boersema PJ, et al. 2017. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science*. 355:eaai7825. 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*. 9:1120–1129. 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:eaav1848. doi:10.1126/sciadv.aav1848.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. 351:652–654. doi:10.1038/351652a0.
- Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*. 52:935–949. 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*. 16:275–284. doi:10.1038/nrg3908.
- Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, et al. 2021. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. 30:187–200. doi:10.1002/pro.3978.
- Peter J, Chiara MD, Friedrich A, Yue J-X, Pflieger D, et al. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*. 556:339–344. doi:10.1038/s41586-018-0030-5.
- Pierleoni A, Martelli PL, Fariselli P, Casadio R. 2007. eSLDB: eukaryotic subcellular localization database. *Nucleic Acids Res*. 35:D208–D212. doi:10.1093/nar/gkl775.
- Salvadó Z, Arroyo-López FN, Guillamón JM, Salazar G, Querol A, et al. 2011. Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl Environ Microbiol*. 77:2292–2302. doi:10.1128/AEM.01861-10.

- Steenwyk JL, Opulente DA, Kominek J, Shen X-X, Zhou X, et al. 2019. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. *PloS Biol.* 17:e3000255. doi:10.1371/journal.pbio.3000255.
- Stern DL. 2014. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizyosity test. *Trends Genet.* 30:547–554. 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.* 4:521–525. doi:10.1016/S1567-1356(03)00171-5.
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, et al. 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49:D605–D612. doi:10.1093/nar/gkaa1074.
- Tam V, Patel N, Turcotte M, Bossé Y, Paré G, et al. 2019. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 20:467–484. 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.* 324:659–662. doi:10.1126/science.1169766.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 25:1189–1191. doi:10.1093/bioinformatics/btp033.
- Weiss CV, Brem RB. 2019. Dissecting trait variation across species barriers. *Trends Ecol Evol.* 34:1131–1136. doi:10.1016/j.tree.2019.07.013.
- Weiss CV, Roop JI, Hackley RK, Chuong JN, Grigoriev IV, et al. 2018. Genetic dissection of interspecific differences in yeast thermotolerance. *Nat Genet.* 50:1501–1504. doi:10.1038/s41588-018-0243-4.
- Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, et al. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio.* 6:e00306–e00315. doi:10.1128/mBio.00306-15.
- Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science.* 285:901–906. doi:10.1126/science.285.5429.901.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591. doi:10.1093/molbev/msm088.

Communicating editor: J. Berman