The polygenic basis of an ancient divergence in yeast thermotolerance

Carly V. Weiss[1], Jeremy I. Roop[1,2], Rylee Hackley[1,3,4], Julie Chuong[3], Igor V. Grigoriev[1,5], Adam P. Arkin[6], Jeffrey M. Skerker[6], Rachel B. Brem[1,3]*

[1]Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA.

[2]Current address: Fred Hutchinson Cancer Research Center, Seattle, WA.

[3]Buck Institute for Research on Aging, Novato, CA.

[4]Current address: Duke University Program in Genetics & Genomics, Duke University, Durham, NC.

[5]US Department of Energy Joint Genome Institute, Walnut Creek, CA.

[6]Department of Bioengineering, UC Berkeley, Berkeley, CA and Lawrence Berkeley National Laboratory, Berkeley, CA.

*Correspondence to: rbrem@buckinstitute.org.

Some of the most unique and compelling survival strategies in the natural world evolved long ago, and are fixed in now-isolated species. Molecular insight into these adaptations has been limited, as classic experimental genetics has focused on the interfertile individuals of a population[1]. Here we dissect a complex thermotolerance difference between yeast species that diverged millions of years ago. Using a new mapping approach that screens mutants in a sterile interspecific hybrid, we identified eight genes that underlie the growth advantage of *Saccharomyces cerevisiae* over its sister species *S. paradoxus* at high temperature. All eight encode housekeeping factors with no known direct function in heat-shock or stress response. Pro-thermotolerance alleles at these mapped loci were required for the adaptive trait in *S. cerevisiae* and sufficient for its partial reconstruction in *S. paradoxus*. Together, our data reveal the genetic mechanism by which *S. cerevisiae* acquired its high-temperature growth advantage in the distant past. And our study lays the groundwork for the mapping of genotype to phenotype in clades of sister species across Eukarya.

Understanding how organisms acquire traits in the wild is a central goal of evolutionary genetics. Given enough time, a phenotype may be refined by evolutionary processes quite different from those that happen early in adaptation[2,3]. Reconstructing these events is a particular challenge when the underlying alleles are fixed in modern-day species. Landmark studies have implicated a single variant locus in a given ancient trait[4,5], but the polygenic mechanisms of species adaptations have remained a mystery.

At high temperature, the yeast *Saccharomyces cerevisiae* grows qualitatively better than other members of its clade[6-8], including its closest relative, *S. paradoxus*, from which it diverged ~5 million years ago[9]. In culture at 39°C, *S. cerevisiae* doubled faster than *S. paradoxus* and accumulated more biomass over a timecourse, a compound trait that we call thermotolerance;

the magnitude of differences in thermotolerance between species far exceeded that of strain variation within each species (Figure 1). No such effect was detectable at 28°C (Figure S1). The failure by *S. paradoxus* to grow to high density at 39°C was, at least in part, a product of reduced survival, as its viability was low starting at quite early times in the culture, relative to that of *S. cerevisiae* (Figure S2).

We set out to dissect the genetic basis of the *S. cerevisiae* thermotolerance phenotype, using a genomic implementation of the reciprocal hemizygote test[10,11] (Figure 2A). We mated *S. cerevisiae* strain DBVPG1373, a soil isolate from the Netherlands, with *S. paradoxus* strain Z1, an English oak tree isolate. In the resulting sterile hybrid we generated hemizygote mutants using a plasmid-borne, selectable PiggyBac transposon system[12]. We cultured the pool of mutants in bulk for ~7 generations at 39°C and 28°C. From cells in each culture we sequenced the genomic context of transposon-containing DNAs[13], as a readout of the genotypes of mutant hemizygote clones present in the selected sample and their abundances. These sequencing data revealed 4888 genes for which transposon mutant clones could be detected in both species' alleles in the hybrid (Figure S3), with transposon insertions distributed in a largely unbiased manner across the genome (Figure S4). For a given gene, we tabulated abundances of mutants whose transposon insertion fell in the *S. cerevisiae* allele of the hybrid, after high-temperature selection relative to the 28°C control, and we compared them to the abundance distribution of mutants in the *S. paradoxus* allele (Figure 2A). Any difference in fitness between these reciprocal hemizygote cohorts can be ascribed to variants between the retained alleles at the respective locus; we refer to the comparison as reciprocal hemizygosity analysis via sequencing (RH-seq). Incorporating this approach with a quality-control pipeline (Figure S3), in a survey of 3416 high-coverage genes we identified 8 top-scoring hits (false discovery rate 0.01; Figure 2B). At each such gene hit, disruption of the *S. cerevisiae* allele in the hybrid was associated with low clone abundance after selection at 39°C relative to 28°C (Figure 2B),

reflecting a requirement for the *S. cerevisiae* allele for thermotolerance. None of the genes mapped by RH-seq had a known role in the yeast heat shock or stress responses. All were annotated as housekeeping factors: *ESP1, DYN1, MYO1, CEP3, APC1,* and *SCC2* in chromosome segregation and cytokinesis, and *AFG2* and *TAF2* in transcription/translation.

We sought to verify, on a single-gene basis, the impact on thermotolerance of allelic variation at genes that emerged from RH-seq. For each of the eight hit genes, we engineered hybrids harboring complete deletions of each species' allele in turn, and we assayed their growth at 28°C and 39°C. At each gene, disrupting the copy from *S. cerevisiae* weakened the thermotolerance of the hybrid (Figure 2B), with little impact on growth at 28°C (Figure S5), consistent with fitness inferences from RH-seq. And at each RH-seq hit gene, the *S. paradoxus* allele made no contribution to the phenotype of the hybrid, since deleting it had no effect (Figure 2B and Figure S5). We conclude that the loci emerging from our RH-seq analysis represent bona fide determinants of thermotolerance in the hybrid.

We expected that our RH-seq hits, though mapped by virtue of effects in the hybrid, could also explain thermotolerance differences between purebred species. As a test of this notion, for each RH-seq hit in turn, we replaced the endogenous allele in each purebred species with the allele from the other. Growth assays of these transgenics established the *S. cerevisiae* allele of each locus as necessary or sufficient for biomass accumulation at 39°C, or both: thermotolerance in the *S. cerevisiae* background was compromised by *S. paradoxus* alleles at 7 of the 8 genes and, in *S. paradoxus*, improved by *S. cerevisiae* alleles at 6 of 8 loci (Figure 3). Allele replacements had little effect on growth at 28°C (Figure S6). Thus, the loci mapped by RH-seq in an interspecies hybrid contribute causally to thermotolerance in purebred backgrounds. Phylogenetic inference among Saccharomycetes made clear that RH-seq hit genes have

evolved rapidly along the branch leading to *S. cerevisiae* (resampling $p = 0.01$), consistent with the likely origin of thermotolerance as a derived trait in this species.

In this work, we have used RH-seq to elucidate the genetic basis of differences in thermotolerance between long-diverged yeasts. Our results suggest that, during its split from *S. paradoxus*, *S. cerevisiae* acquired thermotolerance in part by tuning chromosome segregation and other cell growth functions. Plausibly, variants at many of our mapped loci could act to improve stability and/or function of the cell division machinery under heat stress[14]. These insights stand in contrast to traditional mapping analyses within *S. cerevisiae*, which have focused on rare thermotolerance losses in strains under relaxed selection[10,15]. Indeed, RH-seq is well-suited to the study of the radiation of species complexes, when each species has developed and refined adaptive traits over long timescales[16]. Given successful mating between any species pair, the sterility of the resulting hybrid need not be a roadblock in the mapping of genotype to phenotype.

Figure 1

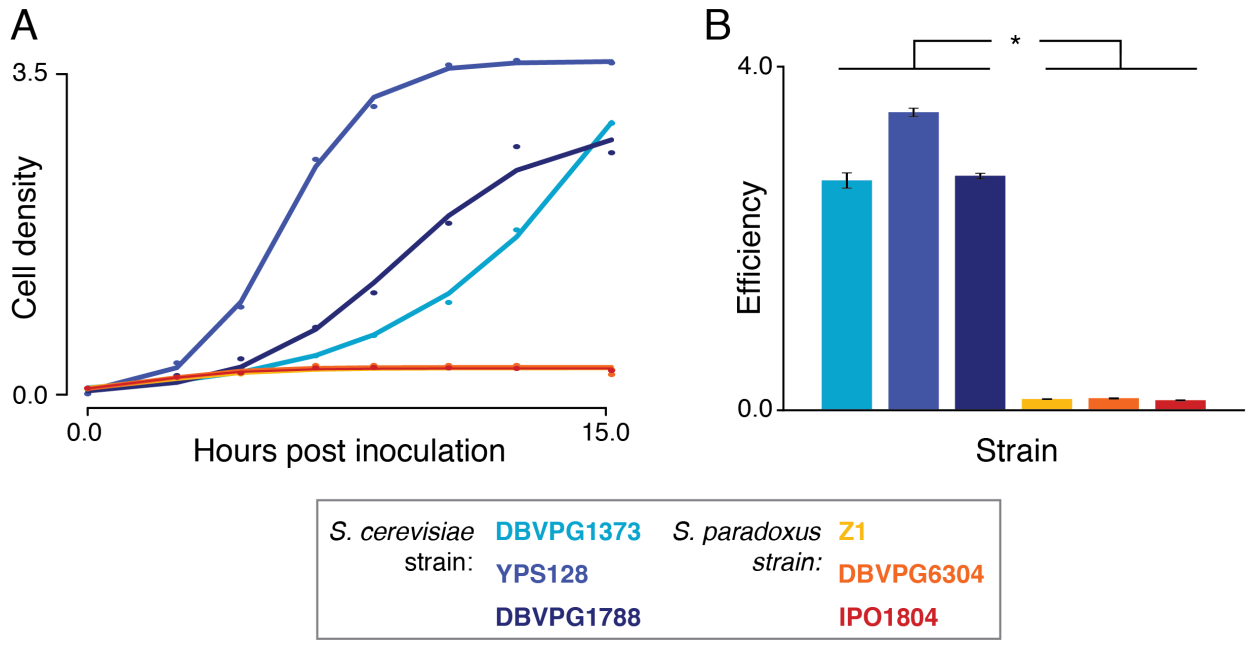

Figure 1. *S. cerevisiae* grows better at high temperature than *S. paradoxus.*

A, Each point reports cell density (OD_{600}) of the indicated wild isolate of *S. cerevisiae* (blue) or *S. paradoxus* (orange) over a timecourse of growth at 39°C. Each solid line shows a logistic population growth model fit to the respective cell density measurements. **B**, Each bar reports mean efficiency (n = 4) of the indicated strain at 39°C, defined as the difference between cell density at 24 hours of growth and that at the time of inoculation. Efficiencies across strains were significantly different between the species (*, p = 3.78×10^{-11}). Error bars report standard deviation.

Figure 2

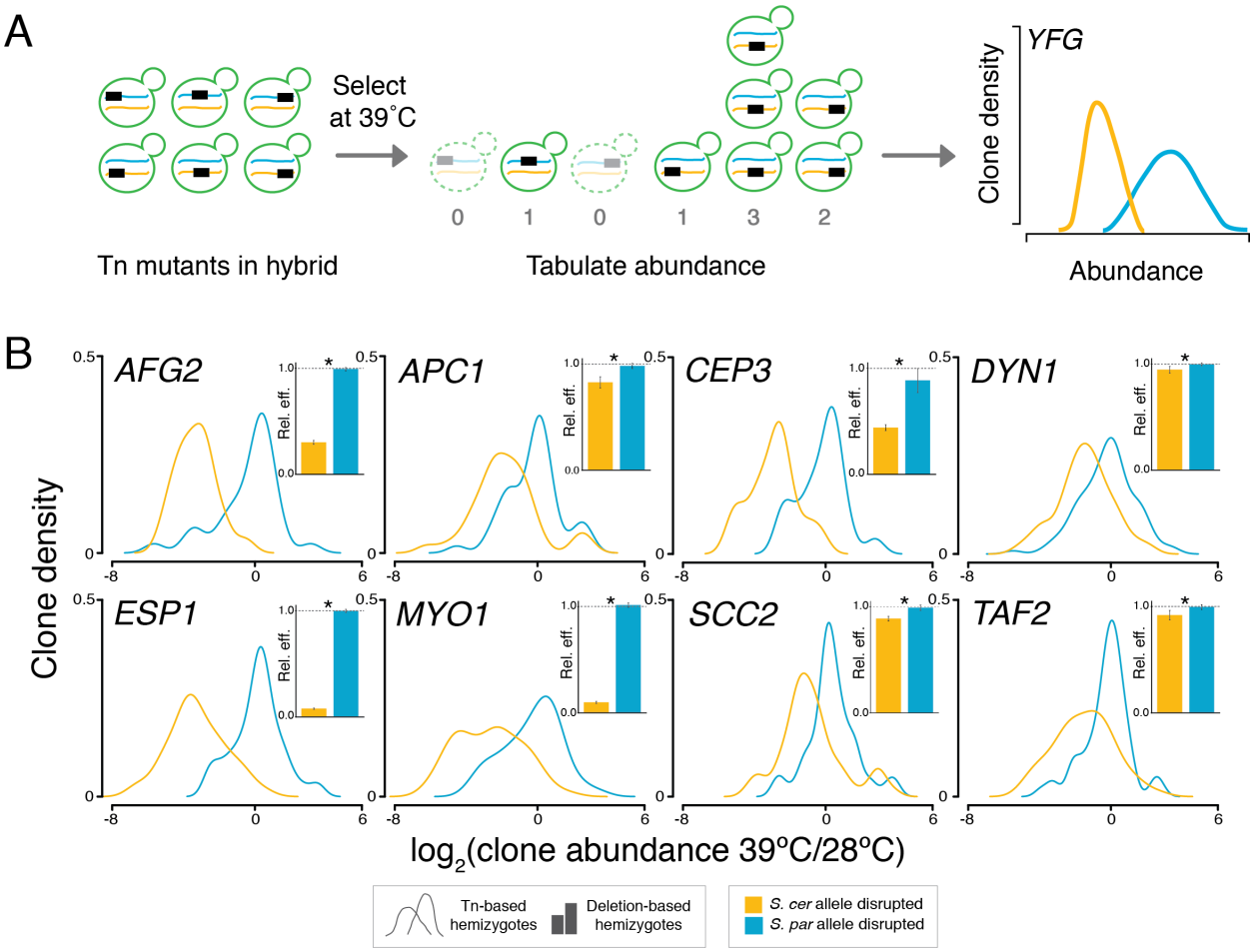

Figure 2. Mapping thermotolerance by RH-seq.

A, A transposon (rectangle) disrupts the allele from *S. cerevisiae* (blue) or *S. paradoxus* (orange) of a gene (*YFG*) in an interspecific hybrid (green). Clones lacking the pro-thermotolerance allele grow poorly at 39°C (dashed outlines), as measured by sequencing and reported in smoothed histograms (traces, colored to indicate the species' allele that is not disrupted). **B**, Each panel reports results from one RH-seq hit locus. In the main figure of a given panel, the *x*-axis reports the $\log_2$ of abundance, measured by RH-seq after selection at 39°C, of a clone harboring a transposon insertion in the indicated species' allele, relative to 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. In insets, each bar reports the relative efficiency: mean growth efficiency at 39°C (n = 8-12) of the indicated targeted-deletion hemizygote measured in liquid culture assays, normalized to the analogous quantity for the wild-type hybrid. *, $p \leq 0.002$; error bars report standard deviation.

Figure 3

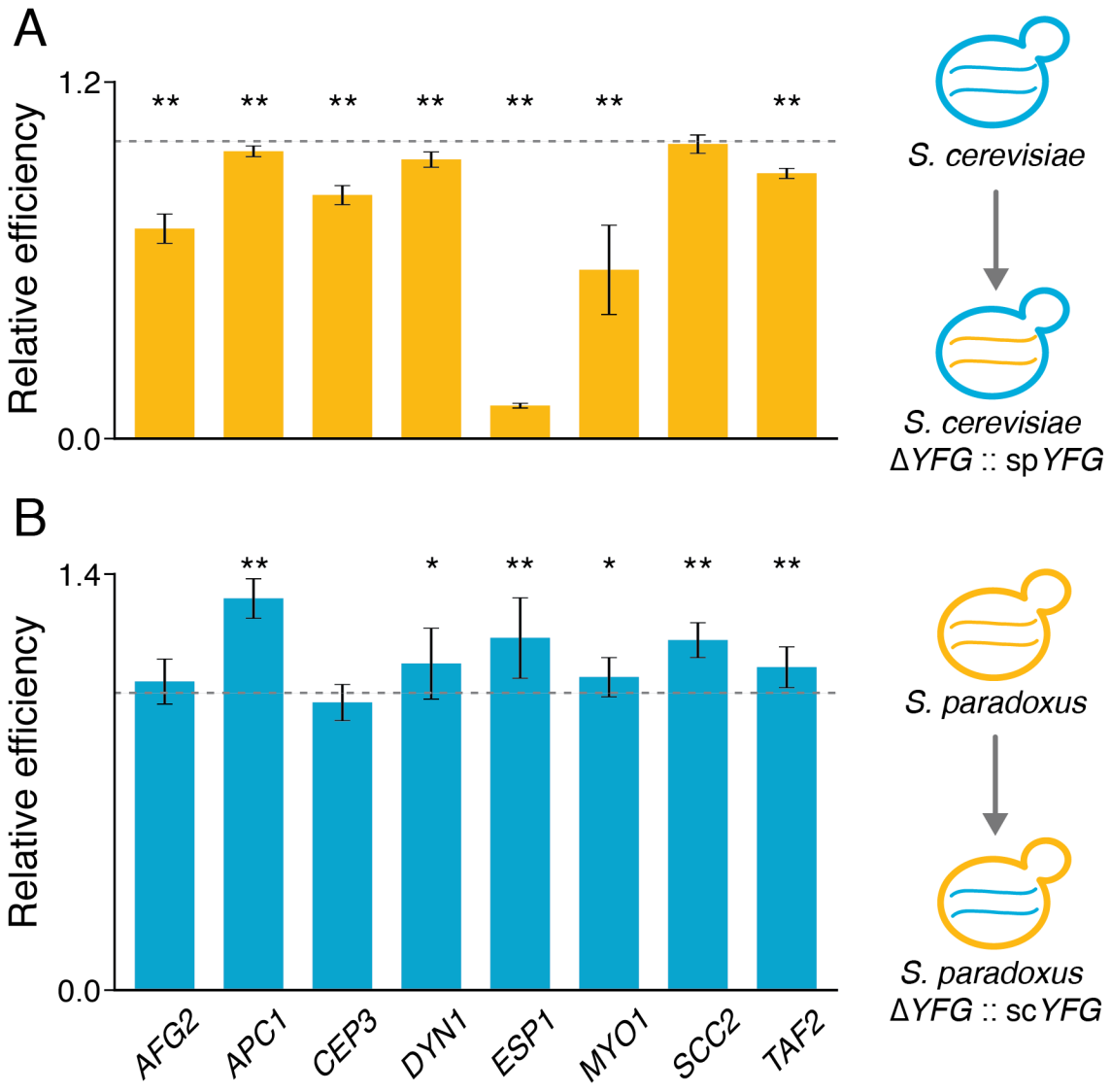

Figure 3. *S. cerevisiae* thermotolerance alleles are necessary and sufficient for growth at high temperature.

A, Each bar reports mean growth efficiency at 39°C, measured in liquid culture assays (*n* = 6-12), of an *S. cerevisiae* strain harboring the *S. paradoxus* allele at the indicated RH-seq hit locus, relative to the analogous quantity for wild-type *S. cerevisiae*. **B**, Data are as in **A**, except that each bar reports results from a *S. paradoxus* strain harboring the *S. cerevisiae* allele at the

indicated locus, normalized to wild-type *S. paradoxus*. *, $p \leq 0.036$; **, $p \leq 0.001$. Error bars report standard deviation.

Materials and Methods

Strains

Strains used in this study are listed in Supplementary Table 1. Homozygous diploid strains of *S. cerevisiae* and *S. paradoxus* used as parents of the interspecific hybrid, and as the backgrounds for allele-swap experiments, were homothallic DBVPG1373 and Z1, respectively. In the case of the hybrid parents, each strain was rendered homozgyous null for *URA3* via homologous recombination with a HYGMX cassette, then sporulated; a given mated spore from a dissected tetrad was grown up into a diploid that was homozygous null at *URA3* and tested for the presence of both genomes by PCR with species-specific primers.

piggyBac transposon machinery

For untargeted, genome-scale construction of reciprocal hemizygotes in the *S. cerevisiae* x *S. paradoxus* hybrid, we adapted methods for piggyBac transposon mutagenesis[12] to develop a system in which the transposon machinery was borne on a selectable and counter-selectable plasmid lacking a centromere. We constructed this plasmid (final identifier pJR487) in three steps. In step I we cloned the piggyBac transposase enzyme gene driven by the *S. cerevisiae TDH3* promoter (from plasmid p3E1.2, a gift from Malcolm Fraser, Notre Dame) into plasmid pJED104 (which contains *URA3*, an ARS, and the *CEN6* locus, and was a gift from John Dueber, UC Berkeley). For this cloning, the amplification used a forward and reverse primer containing a BamHI and XhoI site, respectively, that upon restriction digest yielded sticky ends for ligation to recipient BamHI and XhoI sites in digested pJED104. We used the resulting plasmid as input into step II, removal of the *CEN6* sequence: we first amplified the entire

plasmid with primers that initiated outside of *CEN6* and were directed away from it, and contained reciprocally complementary NheI sites; sticky ends of the linear PCR product were then ligated together for re-circularization. We used the resulting plasmid as input into step III, the cloning in of a construct comprised of the KANMX cassette flanked by long terminal arms (328bp and 361bp) from the piggyBac transposon. We first amplified KANMX from pUG6[17] and each transposon arm from p3E1.2, using primers that contained overlapping sequence on the fragment ends that would ultimately be the interior of the construct, and XbaI sites on the fragment ends that would ultimately be the 5' and 3'-most ends of the construct. We stitched the three fragments together by overlap extension PCR, digested the resulting construct and the plasmid from step II with XbaI, and annealed sticky ends of the two to yield the final pJR487 plasmid.

Untargeted hemizygote construction via transposon mutagenesis

For mutagenesis, pJR487 was gigaprepped using a column kit (Zymo Research) to generate ~11 mg plasmid. To prepare for transformation, JR507 (the *S. cerevisiae* x *S. paradoxus* hybrid) was streaked from a -80°C freezer stock onto a yeast peptone dextrose (YPD, 1% yeast extract [BD], 2% yeast peptone [BD], 2% D-glucose [Sigma]) agar plate and incubated for 2 days at 26°C. A single colony was inoculated into 100 mL YPD and shaken at 28°C, 200rpm for ~24 hours. The next day, we transferred cells from this pre-culture, and YPD, to each of four 1 L flasks at the volumes required to attain an optical density at 600 nm (OD_{600}) of 0.2 in 500 mL each. We cultured each for 6 hours at 28°C with shaking at 200rpm. Two of these cultures were combined into 1 L of culture and two into a separate 1 L, and each such culture was subjected to transformation (for a total of two transformations) as follows. The 1 L was split into twenty 50-mL conical tubes. Each aliquot was centrifuged and washed with water and then with 0.1 M lithium acetate (LiOAc, Sigma) mixed with 1X Tris-EDTA buffer (10 mM Tris-HCl and 1.0 mM

EDTA); after spin-down, to each tube was added a solution of 0.269 mg of pJR487 mixed 5:1 by volume with salmon sperm DNA (Invitrogen), and then to each was added 3 mL of 39.52% polyethylene glycol, 0.12M LiOAc and 1.2X Tris-EDTA buffer (12 mM Tris-HCl and 1.2 mM EDTA). Tubes were rested for 10 minutes at room temperature, then heat-shocked in a water bath at 39°C for 26 minutes. Cells from all 20 tubes were then combined. We transferred cells from this post-transformation culture, and YPD, to each of three 1 L flasks at the volumes required to attain an OD_{600} of ~0.35-4 in 500 mL. Each such culture was recovered by shaking at 28°C and 200 rpm for 2 hours. G418 (Geneticin, Gibco) was added to each at a concentration of 300 µg/mL to select for those cells which had taken up the plasmid, and cultures were incubated with 200 rpm shaking at 28°C for two days until each reached an OD_{600} of ~2.3. All six such selected cultures across the two transformations were combined. We transferred cells from this combined culture, and YPD + G418 (300 ug/mL), to each of two 1 L flasks at the volumes required to attain an OD_{600} of 0.2 in 500 mL each. We cultured each flask at 28°C and 200 rpm shaking overnight until reaching an OD_{600} of 2.18 and combined the two cultures again to yield one culture. To cure transformants of the pJR487 URA+ plasmid, we spun down a volume of this master culture and resuspended in water with the volume required to attain a cell density of 1.85 OD_{600} units/mL. 12 mL of this resuspension were plated (1 mL per 24.1cm x 24.1cm plate) onto plates containing complete synthetic media with 5-fluoroorotic acid (5-FOA) [0.2% drop-out amino acid mix without uracil or yeast nitrogen base (YNB) (US Biological), 0.005% uracil (Sigma), 2% D-glucose (Sigma), 0.67% YNB without amino acids (Difco), 0.075% 5-FOA (Zymo Research)]. After incubation at 28°C to enable colony growth, colonies were scraped off all 12 plates and combined into water at the volume required to attain 40 OD_{600} units per 900 µL, yielding the final transposon mutant hemizygote pool. This was aliquoted into 1 mL volumes with 10% DMSO and frozen at -80°C.

Thermotolerance phenotyping via selection of the hemizygote pool

One aliquot of the pool of transposon mutant hemizygotes in the JR507 *S. cerevisiae* x *S. paradoxus* background was thawed and inoculated into 150 mL of YPD in a 250 mL flask, and cultured for 7.25 hours at 28°C, with shaking at 200 rpm. We used this timepoint as time zero of our thermotolerance experiment, and took four aliquots of 6.43 mL (7 OD units) as technical replicates for sequencing of transposon insertion positions (see below). 9.19 mL of the remaining culture was back-diluted to an OD_{600} of 0.02 in a total of 500 mL YPD in each of six 2L glass flasks for cultures that we call selections; three were grown at 28°C and three at 39°C (shaking at 200 rpm) until an OD_{600} of 1.9-2.12 was reached, corresponding to about 6.5 doublings in each case. Four cell pellets of 7 OD_{600} units each were harvested from each of these biological replicate flasks, for sequencing as technical replicates (see below). In total, 28 pellets were subjected to sequencing: 4 technical replicates from the time-zero culture; 3 biological replicates, 4 technical replicates each, from the 28°C selection; and 3 biological replicates, 4 technical replicates each, from the 39°C selection.

Tn-seq library construction

To determine the abundance of transposon mutant hemizygote clones after selection, we first sequenced transposon (Tn) insertions as follows. Each cell pellet from a time zero or selection sample (see above) 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) pre-warmed to 65°C and its concentration was quantified using a Qubit 3.0 fluorometer. Illumina Tn-seq library construction was as described[18]. Briefly, gDNA was sonicated and ligated with common adapters, and for each fragment deriving from a Tn insertion in the genome, a sequence containing a portion of the transposon and a portion of its genomic context (the Tn-genome junction) was amplified using one primer homologous to a region in the

transposon, and another primer homologous to a region in the adapter. The indexed adapter-specific primer was CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT, where the six N's are a unique index used for multiplexing multiple libraries onto the same Hiseq sequencing lane, and the transposon specific primer was ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*TNNNNNNAGCAATATTTCAAGAATGCATGCGTCAAT, where the asterisk refers to a phosphorothioate modification and the six N's are random nucleotides. Amplification used Jumpstart polymerase (Sigma) and the following cycling protocol: 94°C-2 min, [94°C-30 sec, 65°C-20 sec, 72°C-30 sec] X 25, 72°C-10 min. Sequencing of single-end reads of 150 bp was done over eight lanes on a HiSeq 2500 at the Joint Genome Institute (Walnut Creek, CA). Reads sequenced per library are reported in Supplementary Table 2.

Tn-seq read-mapping and data analysis

For analysis of data from the sequencing of Tn insertion sites in pools of hemizygotes, we first searched each read for a string corresponding to the last 20 base pairs of the left arm of the piggyBac transposon sequence, allowing up to two mismatches. For each Tn-containing read, we then identified the genomic location of the sequence immediately downstream of the Tn insertion site, which we call the genomic context of the insertion, by mapping with BLAT (minimum sequence identity = 95, tile size = 12) against a hybrid reference genome made by concatenating amended *S. cerevisiae* and *S. paradoxus* genomes (see below). These genomic-context sequence fragments were of variable length; any case in which the sequence was shorter than 50 base pairs was eliminated from further analysis, as was any case in which a genomic-context sequence mapped to more than one location in the hybrid reference. The resulting data set thus comprised reads containing genomic-context sequences specifically

mapping to a single location in either *S. cerevisiae* or *S. paradoxus*, which we call usable reads. For a given library, given a cohort of usable reads whose genomic-context sequence mapped to the same genomic location, we inferred that these reads originated from clones of a single mutant with the Tn inserted at the respective site, which we call an insertion. In cases where the genomic-context sequences from reads in a given library mapped to positions within 3 bases of each other, we inferred that these all originated from the same mutant genotype and combined them, assigning to them the position corresponding to the single location to which the most reads mapped among those combined. For a given insertion thus defined, we considered the number of associated reads n_{insert} as a measure proportional to the abundance of the insertion clone in the cell pellet whose gDNA was sequenced. To enable comparison of these abundances across samples, we tabulated the total number of usable reads n_{pellet} from each cell pellet, took the average of this quantity across pellets, $<n_{pellet}>$, and multiplied each n_{insert} by $<n_{pellet}>/ n_{pellet}$ to yield a_{insert}, the final normalized estimate of the abundance of the insertion clone in the respective pellet. For any insertions that were not detected in a given pellet's library ($n_{insert} = 0$) but detectable in another library of the data set, we assigned $n_{insert} = 1$.

We evaluated, from the mapped genomic-context sequence of each insertion, whether it fell into a gene according to the *S. cerevisiae* and *S. paradoxus* genome annotations[9,19], and we retained for further analysis only those insertions that fell into genes that were syntenic in the two species. For each such insertion, for each biological replicate corresponding to a selection culture (at 28°C or 39°C), we averaged the normalized abundances a_{insert} across technical replicates, yielding a single abundance estimate $<a_{insert}>_{technical}$ for the biological replicate. We then calculated the mean of the latter quantities across all biological replicates of the selection, to yield a final abundance estimate for the insertion in this selection, $<a_{insert}>_{total}$. Likewise, for each insertion and selection experiment we calculated $CV_{insert,total}$, the coefficient of variation of $<a_{insert}>_{technical}$ values across biological replicates.

To use Tn-seq data in reciprocal hemizygosity tests, we considered for analysis only genes annotated with the same (orthologous) gene name in the *S. cerevisiae* and *S. paradoxus* reference genomes. For each insertion, we divided the $\langle a_{insert} \rangle_{total}$ value from the 39°C selection by the analogous quantity from the 28°C selection and took the $\log_2$ of this ratio, which we consider to reflect thermotolerance as measured by RH-seq. For each gene in turn, we used a two-tailed Mann-Whitney U test to compare thermotolerance measured by RH-seq between the set of insertions falling into the *S. cerevisiae* alleles of the gene, against the analogous quantity from the set of insertions falling into the *S. paradoxus* allele of the gene, and we corrected for multiple testing using the Benjamini-Hochberg method.

We tabulated the number of inserts and genes used as input into the reciprocal hemizygote test, and the number of top-scoring genes emerging from these tests, under each of a range of possible thresholds for coverage and measurement noise parameter values (Supplementary Figure 3). We used in the final analysis the parameter-value set yielding the most extensive coverage and the most high-significance hits: this corresponded to insertions whose abundances had, in the data from at least one of the two selections (at 28°C or 39°C), $CV_{insert,total} \leq 1.5$ and $\langle a_{insert} \rangle_{total} \geq 1.1$, and genes for which this high-confidence insertion data set contained at least 5 insertions in each species' allele. This final data set comprised 110,678 high-quality insertions (Supplementary Table 3) in 3416 genes (Supplementary Table 4).

Amended reference genome construction

We generated reference genomes for *S. cerevisiae* strain DBVPG1373 and *S. paradoxus* strain Z1 as follows. Raw genome sequencing reads for each strain were downloaded from the SGRP2 database (ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/SGRP2/input/strains). Reads were

aligned using bowtie2[20] with default options; DBVPG1373 reads were aligned to version R64.2.1 of the reference sequence of the *S. cerevisiae* type strain S288C (Genbank Assembly Accession GCA_000146045.2), and Z1 reads were aligned to the *S. paradoxus* strain CBS432 reference sequence[21]. Single nucleotide variants (SNPs) were called using a pipeline of samtools[22], bcftools and bgzip, and were filtered for a quality score (QUAL) of >20 and a combined depth (DP) of >5 and either <65 (*S. cerevisiae*) or <255 (*S. paradoxus*). We then amended each reference genome with the respective filtered SNPs: we replaced the S288C allele with that of DBVPG1373 at each filtered SNP using bcftools' consensus command with default options (42,983 base pairs total), and amendment of the CBS432 sequence was carried out analogously using Z1 alleles (15,126 base pairs total).

Targeted-deletion hemizygote construction by homologous recombination

A given targeted hemizygote for each RH-seq hit gene except *TAF2* was generated in the *S. cerevisiae* x *S. paradoxus* hybrid (JR507) by knocking out the allele of the gene from one species via homologous recombination with KANMX as described[23] with 70 base pairs of homology on the 5' and 3' ends of the cassette; checking was via diagnostic PCR. Two or more independent transformants were isolated and phenotyped for each hemizygote genotype (Supplementary Table 1).

Construction of allele replacement and targeted hemizgyote strains with Cas9

At each RH-seq hit gene, we constructed strains in the wild-type homozygous diploid *S. cerevisiae* (DBVPG1373) in which both copies of the endogenous allele were replaced by the allele from *S. paradoxus* (Z1), and likewise for *S. cerevisiae* alleles into *S. paradoxus*. We call each such strain an allele-replacement strain, and each was constructed using a dual-guide

Cas9 transgenesis method[24] in which a linear PCR fragment from the donor species is incorporated into the recipient genome by homology-directed repair of two chromosomal double-strand breaks induced by Cas9. Briefly, for each allele of each gene, we designed two guide RNAs for double-strand breaks by Cas9: one guide targeted a position ~1000 base pairs 5' to the coding start or at the 3' end of the closest upstream gene, whichever was closer, and the other guide targeted the region of the coding stop. The precise cut site of each was chosen to contain an NGG immediately downstream of variants between the *S. cerevisiae* and *S. paradoxus* strains, to avoid re-cutting of the donor allele by Cas9 after it had been introduced into the recipient strain. We cloned the two guide RNAs, a KANMX cassette, and the gene encoding the *S. pyogenes* Cas9 protein into a single plasmid as described[24]. The resulting plasmid was propagated in DH5a *E. coli* and miniprepped with a column miniprep kit (Qiagen). Separately, to generate the fragment to be used as the donor for DNA repair after Cas9 cutting, we PCR-amplified the respective region from the donor strain, with primers whose 5'-most 70 base pairs were homologous to the recipient and whose 3'-most 20 base pairs were homologous to the donor. The homology region at the 5' end of the gene ended at most 31 base pairs upstream of the 5' cut site, and the homology region at the 3' end of the gene started at most 33 base pairs downstream of the 3' cut site. The donor fragment product was purified with a column kit (Qiagen) and ethanol-precipitated. We then simultaneously transformed, using the lithium acetate method, the donor fragment and dual-guide Cas9 plasmid into the recipient strain, using donor:acceptor ratios of 1:0.353 to 1:5, with 1-10ug of plasmid. In this transformation, heat shock was for 20 minutes at 42°C in transformations of *S. cerevisiae* (DBVPG1373), and ten minutes at 37°C for *S. paradoxus* (Z1). Transformants were plated on YPD+G418 (300 µg/mL) to select for cells that retained the plasmid. From this selection we patched single colonies onto YPD without G418, under the expectation that by the time a lawn came up for each patch, its cells would have lost the Cas9 plasmid. Each such strain was Sanger-sequenced at the junctions of the recipient and donor sequence. Positive patches were

streaked to single colonies on YPD plates, and cells from each such colony were used to inoculate a patch on a YPD plate and, separately, to inoculate a patch on a YPD+G418 plate. Those colonies whose patches grew on the former but not on the latter were inferred to be cured of the plasmid and stored at -80°C. For all genes except *DYN1*, two such strains from each transformation were retained for thermotolerance assays and underwent Sanger sequencing of the entire locus to determine the exact swapped region (Supplementary Table 1). For *DYN1* allele-replacement in *S. paradoxus*, the Cas9-based strategy yielded a single verified clone in which the *S. paradoxus* allele of *DYN1* was replaced by that of *S. cerevisiae*, and likewise for *DYN1* allele-replacement in *S. cerevisiae*. In each case, we mated the single swap clone to a wild-type of the respective species background, confirmed heterozygosity of the resulting diploid via allele-specific diagnostic PCR at the *DYN1* locus, sporulated, and dissected tetrads, allowing each spore to autodiploidize and grow up as a homozygote; we retained from one such tetrad the two spores that were homozygous at the *DYN1* locus for the swapped allele, as confirmed by sequencing, and stored these at -80°C.

Targeted-deletion hemizygote strains for *TAF2* were generated by knocking out the *S. cerevisiae* or the *S. paradoxus* allele in the interspecific hybrid (JR507) using the above methods for Cas9 cutting and repair, with the following differences. To generate the fragment to be used as the donor for DNA repair after Cas9 cutting, we PCR-amplified the NATMX cassette from pBC713 (a gift from John Dueber, constructed as in [25]) using primers whose 5'-most 70 base pairs were homologous to the recipient and whose 3'-most 20 base pairs were homologous to the cassette. The homology region at the 5' end of the gene ended 22 base pairs outside the 5' cut site (which was upstream of coding start), and the homology region at the 3' end of the gene started 33 base pairs outside the 3' cut site. Positive strains were confirmed by PCR. Two independent transformants were isolated and phenotyped for each genotype (Supplementary Table 1).

Growth assays

Growth measurements of wild-type SGRP strains.

For the growth timecourse of a given SGRP strain at 28°C, it was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. A single colony was inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm. This culture was back-diluted into YPD at an OD_{600} of ~0.05 and grown for an additional 5.5 hours at 28°C, 200 rpm, until reaching logarithmic phase. We transferred cells from each such pre-culture, and YPD, to each of 11 replicate wells of a 96-well plate, with volumes sufficient to yield a total volume of 150 µL per well at an OD_{600} of 0.02. The plate was covered with a gas-permeable membrane (Sigma) and incubated with orbital shaking in an M200 plate reader (Tecan, Inc.) at 28°C for 24 hours. For curves in Supplementary Figure 1A, measurements for optical density at 595nm (OD_{595}) were taken every 30 minutes and for each timepoint, the average was taken across replicate wells. To subtract background OD_{595} for the resulting curve, we tabulated the mean of the five lowest values from all datapoints, excluding the first two, and subtracted this value from that of each timepoint, setting any negative value to 0. To smooth the resulting curve, we first replaced each timepoint measurement by its average with those of the timepoints immediately before and after it; then, for any timepoint whose measurement was not greater than or equal to the previous one, we set it to be equal to that previous data point. For Supplementary Figure 1B, the efficiency for a given growth curve (from a single well) was calculated as the difference between the OD_{595} measured at the last four smoothed and averaged data points and that of the first four smoothed and averaged data points. Efficiencies from all of the wells from every *S. cerevisia*e strain were combined and compared to efficiencies from all of the wells for every *S. paradoxus* strain in a two-sample two-tailed *t*-test.

For the growth timecourse of a given SGRP strain at 39°C (Figure 1A), we used a large-volume flask growth paradigm to avoid the influence of plate effects on growth measurements at high temperature in the incubated microplate reader, as follows. Each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. A single colony of a given strain was inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm. Each of these cultures was back-diluted into YPD at an OD_{600} of 0.05 and grown for an additional 5.5-7.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. We transferred cells from each such pre-culture, and YPD, to a glass 250 mL flask at the volumes required to attain an OD_{600} of 0.05 in 100 mL YPD, and incubated it at 39°C with shaking at 200 rpm. OD_{600} readings were taken every ~2 hours for ~18 hours. Figure 1A reports representative data from one of three such independent timecourse experiments. For curve fits, we used the getInitial and SSlogis functions in R to estimate starting values for the parameters of the logistic equation, and the nls function to fit the final parameters.

For efficiency measurements of a given SGRP strain at 39°C in the large-volume format (Figure 1B), it was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. Two single colonies of each strain were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate two replicate pre-cultures. Each was back-diluted into YPD at an OD_{600} at 600 nm of 0.05 and grown for an additional 5.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. The two pre-cultures were each again back-diluted into YPD in 1-inch diameter glass tubes with a target OD_{600} of 0.05; the actual OD_{600} of each was measured, after which it was grown at 39°C with shaking at 200rpm for 24 hours, and OD_{600} was measured again. The efficiency for each replicate was calculated as the difference between these final and initial OD_{600} values. The pipeline from inoculation off solid plates through pre-culture, two back-dilutions, and growth at 39°C we refer to as a day's

growth experiment for an SGRP strain. For each day's experiments, we calculated the average efficiency across the replicates of each strain $<e_{strain}>$. We carried out two days' worth of replicate growth experiments for each strain. For a given species, we used the complete cohort of measurements of $<e_{strain}>$ from all strains of each species, across all days, as input into a two-sample, two-tailed *t*-test to evaluate whether the suite of e_{strain} values across strains of *S. cerevisiae* was significantly different from the analogous set of values from *S. paradoxus*.

Growth measurements of targeted-deletion hemizygotes and allele-replacement strains at 28°C.

For efficiency measurements of a given targeted-deletion hemizygote or allele-replacement strain at 28°C (Supplementary Figures 5 and 6), pre-culture and plate reader assays were as for wild-type SGRP strains, except that 6 or more replicate wells were cultured per strain. Two independently isolated targeted-deletion hemizygotes or allele-replacement strains (Supplementary Table 1) were assayed for each genotype. Each timecourse of targeted-deletion hemizygote or allele-replacement strains also included the wild-type hybrid (JR507) or parent (*S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1), respectively, with pre-culture and replication as above. Efficiency for a given growth curve (from a single well) was calculated as the difference between the OD_{600} measured at the last four smoothed and averaged datapoints and that of the first four smoothed and averaged datapoints, with smoothing and averaging as detailed above. For Supplementary Figure 5, relative efficiency for a given well of a given targeted-deletion hemizygote strain at 28°C was tabulated as its efficiency divided by that of the average of all replicate wells of the wild-type hybrid grown in the same experiment. For a given gene, we used the complete cohort of these measurements, from all isogenic hemizygotes, as input into a two-sample, two-tailed *t*-test to evaluate whether the relative efficiency of the strain in which the *S. cerevisiae* allele was knocked out was lower than the analogous quantity from the strain in which the *S. paradoxus* allele was knocked out. In Supplementary Figure 6, allele-

replacement strains for a given gene were analyzed analogously using the respective wild-type parent and with a one-sample, two-tailed *t*-test to evaluate whether the relative efficiency was significantly different from 1.

Growth measurements of targeted-deletion hemizygotes and allele-replacement strains at 39°C.

For efficiency measurements of a given targeted-deletion hemizygote strain at 39°C in the large-volume format (Figure 2B), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. Two single colonies of a given strain were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm. Each such pre-culture at stationary phase, or a log-phase outgrowth of it (which we used in the case of *DYN1* and *TAF2*: the pre-culture and YPD were added at the volumes required to attain an OD_{600} of 0.05 and grown for an additional 5.5 hours at 28°C, shaking at 200 rpm, until the culture reached logarithmic phase) was used to inoculate YPD in 1-inch diameter glass culture tubes with a target cell density corresponding to an OD_{600} of 0.05. The actual OD_{600} of each was measured, after which it was grown at 39°C with shaking at 200 rpm for 24 hours, and OD_{600} was measured again. The efficiency for each such replicate was then calculated as the difference between the final and initial OD_{600} values. The pipeline from inoculation off solid plates through pre-culture, back-dilution, and growth at 39°C we refer to as a day's growth experiment for a targeted-deletion homozygote. In each such experiment, 2-3 independently isolated targeted-deletion hemizygotes of a given gene in each direction were all assayed on the same day, alongside the wild-type hybrid parent (JR507) with replicate structure and methods as above. From each day's experiments, we calculated the average efficiency across the replicates of the wild-type hybrid $\langle e_{hybrid} \rangle$, and we used this quantity to normalize the efficiency $e_{hemizyg}$ measured for each replicate of each hemizygote strain assayed on that day. Thus, the final observable used for analysis for each replicate on a given day was

$e_{hemizyg}/\langle e_{hybrid}\rangle$. We carried out 2-3 days' replicate growth experiments for each gene's hemizygotes. For a given gene, we used the complete cohort of these measurements of $e_{hemizyg}/\langle e_{hybrid}\rangle$, from all days and all isogenic hemizygotes, as input into a two-sample, one-tailed *t*-test to evaluate whether $e_{hemizyg}/\langle e_{hybrid}\rangle$ of the strain in which the *S. cerevisiae* allele was knocked out was lower than the analogous quantity from the strain in which the *S. paradoxus* allele was knocked out.

For growth measurements of a given allele-replacement strain at 39°C in the large-volume format (Figure 3), each strain was streaked from a -80°C freezer stock onto a YPD agar plate and incubated at 26°C for 3 days. Two single colonies of each strain were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to generate two replicate pre-cultures. Each was back-diluted into YPD at an OD_{600} of 0.05 and grown for an additional 5.5 hours at 28°C, shaking at 200 rpm, until reaching logarithmic phase. The two pre-cultures were each again back-diluted into YPD in 1-inch diameter glass tubes with a target OD_{600} of 0.05; the actual OD_{600} of each was measured, after which it was grown at 39°C with shaking at 200rpm for 24 hours, and OD_{600} was measured again. The efficiency for each replicate was calculated as the difference between these final and initial OD_{600} values. The pipeline from inoculation off solid plates through pre-culture, two back-dilutions, and growth at 39°C we refer to as a day's growth experiment for an allele-swap strain. In each such experiment, 2-3 independently isolated allele-swap strains targeting a given gene in a given background were all assayed on the same day, alongside the respective wild-type background strain (*S. cerevisiae* DBVPG1373 or *S. paradoxus* Z1) with replicate structure and methods as above. For each day's experiments, we calculated the average efficiency across the replicates of the wild-type parent $\langle e_{parent}\rangle$, and we used this quantity to normalize the efficiency e_{swap} measured for each replicate assayed on that day of each allele-swap strain in the respective background. Thus, the final measurement used for analysis for each replicate on a given day was $e_{swap}/\langle e_{parent}\rangle$. We

carried out 2-3 days' worth of replicate growth experiments for each gene's allele-swap strains. For a given gene in a given background, we used the complete cohort of measurements of $e_{swap}/\langle e_{parent}\rangle$ from all days and all allele-swap strains as input into a one-sample, one-tailed *t*-test to evaluate whether $e_{swap}/\langle e_{parent}\rangle$ was significantly different from 1. For swaps of the *S. cerevisiae* allele of a given gene into *S. paradoxus*, we tested whether $e_{swap}/\langle e_{parent}\rangle$ was greater than 1 (*i.e.* that the swap strain grew better at 39°C than did its parent), and for swaps of the *S. paradoxus* allele of a given gene into *S. cerevisiae*, we tested whether $e_{swap}/\langle e_{parent}\rangle$ was less than 1 (*i.e.* that the swap strain grew worse at 39°C than its parent).

Testing viability of wild-type parent strains at 39°C.

To test the viability of heat-treated *S. paradoxus* Z1 and *S. cerevisiae* DBVPG1373 (Supplementary Figure 2), each strain was streaked from -80°C freezer stocks onto YPD agar plates and incubated at 26°C for 3 days. 2-3 single colonies of each parent were each inoculated into liquid YPD and grown for 24 hours at 28°C with shaking at 200 rpm to create replicate pre-cultures. After 24 hours, we transferred cells from each pre-culture, and YPD, to each of two tubes at the volumes required to attain an OD_{600} of 0.05 in 11 mL YPD. One tube from each pre-culture was incubated at 28°C and one at 39°C, all with shaking at 200 rpm for 6 hours, which we call treated cultures. OD_{600} of each was measured, and 100 µL of a 5.0×10^{-4} serial dilution of each was plated to YPD (excepting *S. paradoxus* cultures grown at 39°C, which were only serially diluted to 10^{-3}). Plates were incubated at 26°C for 3 days until single colonies appeared. Colonies on each plate were counted, from which we tabulated the colony forming units (CFU) per mL of culture plated. Viability was then calculated for each treated culture as the ratio of CFU/mL to OD_{600}, and we computed the relative viability for the two treated cultures from a given pre-culture (treated at 39°C and 28°C respectively) as the ratio of their viabilities, $([CFU/mL]/OD_{600})_{39}$ / $([CFU/mL]/OD_{600})_{28}$. The pipeline from pre-culture through treatment and

colony counting we refer to as one day's worth of experiments. All values from both species across two days' worth of experiments were used in a two-sample, one-tailed *t*-test to evaluate whether the relative viability of *S. paradoxus* was lower than that of *S. cerevisiae.*

Phylogenetic analysis

We downloaded orthologous protein coding regions for the type strains of *S. cerevisiae*, *S. paradoxus*, and an outgroup, *S. mikatae*, from [9]. For each gene for which ortholog sequences were available in all three species, we aligned the sequences with PRANK[26] utilizing the "-codon" option for codon alignment. These alignments were used as input into the codeml module of PAML[27], which was run assuming no molecular clock and allowing omega values to vary for each branch in the phylogeny. From the resulting inferences, we tabulated the branch length on the *S. cerevisiae* lineage for each gene. To evaluate whether sequence divergence of the eight RH-seq hit genes showed signatures of rapid evolution along the *S. cerevisiae* lineage, we first tabulated $\langle l \rangle_{true}$, the mean of these branch lengths across the eight RH-seq hit genes. We then sampled eight random genes from the set of 3416 genes tested by RH-seq; to account for biases associated with lower rates of divergence among essential genes, the resampled set contained six essential genes and two non-essential genes, mirroring the breakdown of essentiality among the RH-seq hits. Across this random sample we tabulated the mean branch length along the *S. cerevisiae* lineage, $\langle l \rangle_{resample}$. We repeated the resampling 5000 times and used as an empirical *p*-value the proportion of resamples at which $\langle l \rangle_{resample} \geq \langle l \rangle_{true}$.

Data availability

RH-seq data will be deposited in the Sequence Read Archive (SRA) upon publication.

Code availability

Custom Python and R scripts used for RH-seq data analysis are available upon request.

References

1 Flint, J. & Mott, R. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat Rev Genet* **2**, 437-445, doi:10.1038/35076585 (2001).

2 Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature*, doi:10.1038/nature24287 (2017).

3 Savolainen, O., Lascoux, M. & Merila, J. Ecological genomics of local adaptation. *Nat Rev Genet* **14**, 807-820, doi:10.1038/nrg3522 (2013).

4 Nadeau, N. J. & Jiggins, C. D. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet* **26**, 484-492, doi:10.1016/j.tig.2010.08.004 (2010).

5 Wray, G. A. Genomics and the Evolution of Phenotypic Traits. *Annual Review of Ecology, Evolution, and Systematics* **44**, 51-72 (2013).

6 Goncalves, P., Valerio, E., Correia, C., de Almeida, J. M. & Sampaio, J. P. Evidence for divergent evolution of growth temperature preference in sympatric Saccharomyces species. *PLoS One* **6**, e20739, doi:10.1371/journal.pone.0020739 (2011).

7 Salvado, Z. *et al.* Temperature adaptation markedly determines evolution within the genus Saccharomyces. *Appl Environ Microbiol* **77**, 2292-2302, doi:10.1128/AEM.01861-10 (2011).

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

9 Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Bethesda)* **1**, 11-25, doi:10.1534/g3.111.000273 (2011).

10 Steinmetz, L. M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326-330, doi:10.1038/416326a (2002).

11 Stern, D. L. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends Genet* **30**, 547-554, doi:10.1016/j.tig.2014.09.006 (2014).

12 Mitra, R., Fain-Thornton, J. & Craig, N. L. piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J* **27**, 1097-1109, doi:10.1038/emboj.2008.41 (2008).

13 van Opijnen, T., Lazinski, D. W. & Camilli, A. Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. *Curr Protoc Mol Biol* **106**, 7 16 11-24, doi:10.1002/0471142727.mb0716s106 (2014).

14 Leuenberger, P. *et al.* Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, doi:10.1126/science.aai7825 (2017).

15 Parts, L. *et al.* Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res* **21**, 1131-1138, doi:10.1101/gr.116731.110 (2011).

16 Allen Orr, H. The genetics of species differences. *Trends Ecol Evol* **16**, 343-350 (2001).

17 Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**, 2519-2524 (1996).

18 Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**, e00306-00315, doi:10.1128/mBio.00306-15 (2015).

19 Skelly, D. A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* **23**, 1496-1504, doi:10.1101/gr.155762.113 (2013).
20 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
21 Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337-341, doi:10.1038/nature07743 (2009).
22 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
23 Roop, J. I. & Brem, R. B. Rare variants in hypermutable genes underlie common morphology and growth traits in wild Saccharomyces paradoxus. *Genetics* **195**, 513-525, doi:10.1534/genetics.113.155341 (2013).
24 Maurer, M. J. *et al.* Quantitative Trait Loci (QTL)-Guided Metabolic Engineering of a Complex Trait. *ACS Synth Biol* **6**, 566-581, doi:10.1021/acssynbio.6b00264 (2017).
25 Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth Biol* **4**, 975-986, doi:10.1021/sb500366v (2015).
26 Loytynoja, A. & Goldman, N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**, 579, doi:10.1186/1471-2105-11-579 (2010).
27 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).

End Notes

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank Gina Geiselman, Matt Maurer, Jake Kim, Justin Hong, and Anna Flury for technical assistance, and Sam Coradetti, Arjun Sasikumar, Ben Blackman, and Chris Nelson for discussions. This work was supported by R01 GM120430-A1 and by Community Sequencing Project 1460 to RBB at the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility. The work conducted by the latter was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Author Contributions R.B.B. and J.I.R. conceived of the project design; C.V.W., J.I.R., R.H. and J.C. performed experiments; C.V.W. and J.I.R. and analyzed the data; J.M.S. and A.P.A. provided Tn-seq reagents; I.V.G. provided technical assistance with sequencing; R.B.B. and C.V.W wrote the manuscript with input from all authors.

Author Information The authors do not declare any competing financial interests. Correspondence and request for materials should be addressed to R.B.B. (rbrem@buckinstitute.org).