Genetic dissection of interspecific differences in yeast thermotolerance

Carly V. Weiss1, Jeremy I. Roop1,6*, Rylee K. Hackley1,2,7, Julie N. Chuong2, Igor V. Grigoriev8,1,3, Adam P. Arkin4,5, Jeffrey M. Skerker4,5 and Rachel B. Brem1,2*

Some of the most unique and compelling survival strategies in the natural world are fixed in isolated species1. To date, molecular insight into these ancient adaptations has been limited, as classic experimental genetics has focused on interfertile individuals in populations2. Here we use a new mapping approach, which screens mutants in a sterile interspecific hybrid, to identify eight housekeeping genes that underlie the growth advantage of Saccharomyces cerevisiae over its distant relative Saccharomyces paradoxus at high temperature. Prothermotolerance alleles at these mapped loci were required for the adaptive trait in S. cerevisiae and sufficient for its partial reconstruction in S. paradoxus. The emerging picture is one in which S. cerevisiae improved the heat resistance of multiple components of the fundamental growth machinery in response to selective pressure. Our study lays the groundwork for the mapping of genotype to phenotype in clades of sister species across Eukarya.

Geneticists since Mendel have sought to understand how and why wild individuals differ. Studies to this end routinely test for a relationship between genotype and phenotype via linkage or association2. These familiar approaches, though powerful in many contexts, have an important drawback—they can only be applied to interfertile members of the same species. This rules out any case in which an innovation in form or function evolved long ago and is now fixed in a reproductively isolated population.

As organisms undergo selection over long timescales, their traits may be refined by processes quite different from those that happen early in adaptation3,4. We know little about these mechanisms in the wild, expressly because when the resulting lineages become reproductively incompatible, classic statistical-genetic methods cannot be used to analyze them1. To date, the field has advanced largely on the strength of candidate-based studies that implicate a single variant gene in an interspecific trait5,6, with the complete genetic architecture often remaining unknown. Against the backdrop of a few specialized introgression7–10 and molecular-evolution11 techniques available in the field, dissection of complex trait differences between species has remained a key challenge.

Here we develop a new genetic mapping strategy, based on the reciprocal hemizygote test12,13, and use it to identify the determinants of a difference in high-temperature growth between isolated Saccharomyces yeast species. We validate the contributions of the mapped loci to the thermotolerance trait, and we investigate their evolutionary history.

At high temperature, the yeast Saccharomyces cerevisiae grows qualitatively better than other members of its clade14–16, including its closest relative, Saccharomyces paradoxus, from which it diverged ~5 million years ago17. In culture at 39 °C, S. cerevisiae doubled faster than S. paradoxus and accumulated more biomass over a time course, a compound trait that we call thermotolerance. The magnitude of differences in thermotolerance between species far exceeded that of strain variation within each species (Fig. 1), whereas no such effect was detectable at 28 °C (Supplementary Fig. 1). The failure by S. paradoxus to grow to high density at 39 °C was, at least in part, a product of reduced survival relative to that of S. cerevisiae, as cells of the former were largely unable to form colonies after heat treatment (Supplementary Fig. 2). In microscopy experiments, S. paradoxus cells were almost uniformly visible as large-budded dyads after 24 h at 39 °C (Supplementary Fig. 3), suggestive of defects late in the cell cycle as a proximal cause of death18. No such tendency was apparent in S. cerevisiae at high temperature, or in either species at 28 °C (Supplementary Fig. 3).

We set out to dissect the genetic basis of S. cerevisiae thermodurability, using a genomic implementation of the reciprocal hemizygote test12,13 (Fig. 2a). For this purpose, we first mated S. cerevisiae strain DBVPG1373, a soil isolate from the Netherlands, with S. paradoxus strain Z1, an English oak tree isolate. The resulting sterile hybrid had a thermodurability phenotype between those of its purebred parents (Supplementary Fig. 4). In this hybrid background we generated hemizygote mutants using a plasmid-borne, selectable piggyBac transposon system19. We cultured the pool of mutants in bulk for approximately seven generations at 39 °C and, separately, at 28 °C. From cells in each culture we sequenced transposon insertion locations20 as a readout of the genotypes and abundance of mutant hemizygote clones present in the selected sample. In these sequencing data, at each of 4,888 genes we detected transposon mutant clones in both species’ alleles in the hybrid (Supplementary Fig. 5), with transposon insertions distributed in a largely unbiased manner across the genome (Supplementary Fig. 6). For a given gene, we tabulated the 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 (Fig. 2a). Any difference in abundance 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). Integrating this

1Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA. 2Buck Institute for Research on Aging, Novato, CA, USA. 3US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. 4Department of Bioengineering, University of California, Berkeley, CA, USA. 5Lawrence Berkeley National Laboratory, Berkeley, CA, USA. 6Present address: Fred Hutchinson Cancer Research Center, Seattle, WA, USA. 7Present address: Duke University Program in Genetics and Genomics, Duke University, Durham, NC, USA. *e-mail: jeremyianroop@gmail.com; rbrem@buckinstiute.org
mapped by virtue of their impact in the hybrid, could also explain minants of thermotolerance in the hybrid. (R validation paradigm largely paralleled the estimates from RH-seq and Supplementary Fig. 7). Locus effect sizes from this single-gene phenotype of the hybrid, since deleting it had no effect (Fig. 2b at each locus, the gene grew poorly at high temperature (Fig. 2b), with little impact allele at each by targeted deletion of each species’ allele in turn in the hybrid. allelic variation rather than an artifact of our genomic approach. between hemizygotes at a given locus were the consequence of from RH-seq, we first sought to verify that growth differences approach with a quality-control pipeline (Supplementary Fig. 5), in a survey of 3,416 high-coverage genes we identified 8 top-scoring hits (false discovery rate 0.01; Fig. 2b and Supplementary Table 1). At each such locus, disruption of the S. cerevisiae allele in the hybrid was associated with low clone abundance after selection at 39°C relative to at 28°C (Fig. 2b), reflecting a requirement for the S. cerevisiae allele for thermotolerance. All of the genes mapped by RH-seq were annotated as housekeeping factors: ESP1, DYN1, MYO1, CEP3, APC1, and SCC2 function in chromosome segregation and cytokinesis and AFG2 and TAF2 in transcription/translation.

To evaluate the role in thermotolerance of genes that emerged from RH-seq, we first sought to verify that growth differences between hemizygotes at a given locus were the consequence of allelic variation rather than an artifact of our genomic approach. To this end, at each RH-seq hit gene we engineered hemizygotes by targeted deletion of each species’ allele in turn in the hybrid. In growth assays, the strain lacking the S. cerevisiae allele at each gene grew poorly at high temperature (Fig. 2b), with little impact at 28°C (Supplementary Fig. 7), as inferred from RH-seq. Likewise, at each locus, the S. paradoxus allele made no contribution to the phenotype of the hybrid, since deleting it had no effect (Fig. 2b and Supplementary Fig. 7). Locus effect sizes from this single-gene validation paradigm largely paralleled the estimates from RH-seq (R² = 0.74). We conclude that RH-seq hits represent bona fide determinants of thermotolerance in the hybrid.

We expected that variation at our RH-seq hits, though mapped by virtue of their impact in the hybrid, could also explain thermotolerance differences between purebred species. As a test of this notion, for each mapped gene in turn, we replaced the two copies of the endogenous allele in each purebred diploid with the allele from the other species. 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 the S. paradoxus was improved by S. cerevisiae alleles at 6 of 8 loci (Fig. 3). Allele replacements had little impact on growth at 28°C (Supplementary Fig 8). These trends mirrored the direction of locus effects from hemizygotes in the hybrid, though the magnitudes were often different. Most salient were the small effect sizes in S. paradoxus relative to other backgrounds, indicative of strong epistasis in this poorly performing species (Supplementary Fig. 9). Thus, the loci mapped by RH-seq in an interspecies hybrid contribute causally to thermotolerance in purebreds, with effect sizes that depend on the context in which they are interrogated.

Avid growth at high temperature is a defining characteristic of S. cerevisiae as a species, relative to other members of Saccharomyces14–16 (Fig. 1). In principle, the loci mapped by RH-seq could be unique to the genetic architecture of thermotolerance in our focal S. cerevisiae strain, DBVPG1373, or it could be part of a mechanism common to many S. cerevisiae isolates. In support of the latter model, transgenesis experiments showed that a diverse panel of S. cerevisiae isolates harbored alleles conferring modest but significant growth benefits at high temperature, and alleles from multiple S. paradoxus isolates were deleterious (Supplementary Fig. 10a,b). We detected no such impact at 28°C (Supplementary Fig. 10a,b). Similarly, we found elevated sequence divergence from S. paradoxus to be a shared feature of S. cerevisiae strains at the loci mapped by RH-seq (using the absolute divergence measure D), and a local divergence population accumulated divergent, pro-thermotolerance alleles at appreciable density in the loci mapped by RH-seq, consistent with a role in the trait for these genes across the species. Additionally, in the yeast phylogeny, RH-seq hit genes were distinguished by accelerated evolution along the branch leading to S. cerevisiae, as expected if the ancestral program has been conserved among the other species in the clade (Supplementary Figure 10c).

In this work, we have developed the RH-seq method for genome-wide mapping of natural trait variation, and we have used it to elucidate the genetics of thermotolerance in reproductively isolated yeasts. Growth at high temperature is likely a derived character in S. cerevisiae14–16, and the mechanism by which evolution built the trait, after the split from S. paradoxus, has remained unknown. In pursuing the genetics of this putative ancient adaptation, we complement studies of younger, intraspecific variants that erode thermotolerance in the few S. cerevisiae isolates that have lost the trait relatively recently 12,21. We have sought to shed light on more ancient evolutionary events by considering S. paradoxus as a representative of the ancestral state, to which thermotolerant S. cerevisiae can be compared.

Using this approach, we have mapped eight loci at which S. cerevisiae alleles are necessary and sufficient for thermotolerance. As our RH-seq scan did not attain complete genomic coverage, the hits we did find likely represent a lower bound on the complexity of the architecture of the trait. Six of the RH-seq hit genes are essential for growth in standard conditions22, and all eight contribute to fundamental growth processes. ESP1, DYN1, CEP3, APC1, MYO1, and SCC2 mediate mitotic spindle assembly, chromatid cohesion and separation, cytokinesis, and mitotic exit; AFG2 regulates the release of maturation factors from the ribosome; and TAF2 encodes a TFIIID subunit. In each case, our growth experiments in the interspecific hybrid have shown that the S. paradoxus allele acts as a hypomorph at high temperature. Our work leaves unanswered exactly how heat-treated S. paradoxus dies in the absence of these functions, though the cells’ large-budded morphology strongly suggests regulated

Fig. 1 | S. cerevisiae grows better at high temperature than S. paradoxus. a. Each point reports optical density (OD600nm) of the indicated wild isolate of S. cerevisiae (various shades of blue) or S. paradoxus (orange) over a time course 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 cultures) of the indicated strain at 39°C, defined as the difference between cell density at 24 h of growth and that at the time of inoculation. *P = 3.78 × 10⁻⁶; two-sample, two-tailed t-test; individual measurements are reported as circles.
arrest or stochastic failure late in the cell cycle. That said, given that some but not all RH-seq hit loci have roles in mitosis, this is likely only one of the choke points at which S. paradoxus alleles are a liability at high temperature. Assuming that these heat-sensitive alleles also littered the genome of the common ancestor with S. cerevisiae, thermostolerance would have evolved along the S. cerevisiae lineage by resolving each of them, boosting the heat resistance of many housekeeping processes. Such a mechanism would dovetail with the recent finding that, across species, the limiting temperature for cell growth correlates with the melting temperatures of a subset of essential proteins.

These insights into the evolution of a complex yeast trait serve as a proof of concept for RH-seq. To date, the reciprocal hemizygosity test has led to landmark discoveries in a candidate-gene framework, confirming the effects of variation at a given locus identified by other means. Schemes to scale up the test have generated a genome’s worth of hemizygotes from deletion-strain purebreds, which tend to harbor secondary mutations that come through screens as false positives. As such, a key advantage of RH-seq is that we carry out mutagenesis in the hybrid, which ensures coverage of essential genes and obviates the use of mutation-prone null genotypes. Furthermore, any secondary mutations that do arise in a given hemizygote clone, for example during a long competition in the condition of interest, would not have a strong influence on RH-seq mapping, because deep mutagenesis generates many independent clones per gene that are analyzed together. One important caveat of RH-seq, as in single-gene reciprocal hemizygote tests, is the assumption that no epistasis unique to the hybrid will mask the effects of loci underlying a trait difference of interest between the parents. In our case study, the genetic architecture of thermostolerance in the hybrid did bear out as relevant for the purebreds, albeit with locus effect sizes that varied across the backgrounds. More dramatic discrepancies may be particularly likely when the hybrid has a heterotic (that is, extreme) phenotype and is a poor model for the genetics of the parents. The choice of a non-heterotic hybrid in which to pursue RH-seq would be analogous to classical linkage mapping in a cross whose progeny have, on average, phenotypes that are intermediate between those of the parents.

In fact, although we have focused here on ancient divergence, the RH-seq method would be just as applicable to individuals within a species, as a high-resolution alternative to linkage analysis. We thus anticipate that RH-seq will accelerate the mapping of genotype to phenotype in many systems, whether the parents of a cross are closely related or members of a species complex that have been locally adapting for millions of years.

**URLs**

SGRP2 Database, ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/SGRP2/input/strains; Yeast Resource Center, http://www.yeastsc.org/g2p/home.do; Saccharomyces Genome Database, http://www.yeastgenome.org; RH-seq data analysis scripts, https://github.com/weiss19/rh-seq.
Fig. 3 | *S. cerevisiae* thermostolerance 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=8–12 cultures), 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*. In a given panel, the top and bottom dotted lines report the relative efficiency of wild-type *S. cerevisiae* and *S. paradoxus*, respectively. *P*≤0.036; **P**≤0.001, one-sample, one-tailed t-test; individual measurements are reported as circles. See Supplementary Table 1 for exact *P*-values and sample numbers.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0243-4.

Received: 6 November 2017; Accepted: 17 August 2018; Published online: 8 October 2018

References
1. Orr, H. A. The genetics of species differences. *Trends Ecol. Evol.* 16, 343–350 (2001).
2. Flint, J. & Mott, R. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat. Rev. Genet.* 2, 437–445 (2001).
3. 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* 551, 45–50 (2017).
4. Savolainen, O., Lascoux, M. & Merila, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14, 807–820 (2013).
5. Nadeau, N. J. & Jiggins, C. D. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet.* 26, 484–492 (2010).
6. Wray, G. A. Genomics and the evolution of phenotypic traits. *Annu. Rev. Ecol. Evol. Syst.* 44, 51–72 (2013).
7. Masly, J. P. & Presgraves, D. C. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biol.* 5, e243 (2007).
8. Greig, D. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genet.* 3, e21 (2007).
9. Eshed, Y. & Zamir, D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141, 1147–1162 (1995).
10. Lazzarino, S. et al. Genetic mapping of species differences via in vitro crosses in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* 115, 3680–3685 (2018).
11. Roop, J. L., Chang, K. C. & Brem, R. B. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature* 530, 336–339 (2016).
12. Steinmetz, L. M. et al. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 416, 326–330 (2002).
13. Stern, D. L. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends Genet.* 30, 547–554 (2014).
14. 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 (2011).
15. Salvado, Z. et al. Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl. Environ. Microbiol.* 77, 2292–2302 (2011).
16. Sweeney, J. Y., Kuehne, H. & Sniegowski, P. D. Sympatric natural *Saccharomyces cerevisiae* and *S. paradoxus* populations have different thermal growth profiles. *FEMS Yeast Res.* 4, 521–525 (2004).
17. Scannell, D. R. et al. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *Genes Genomes Genet.* 1, 11–25 (2011).
18. Hartwell, L. H. *Saccharomyces cerevisiae* cell cycle. *Bacteriol. Rev.* 38, 164–198 (1974).
19. Mitra, R., Fain-Thornton, J. & Craig, N. L. PiggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.* 27, 1097–1109 (2008).
20. van Opinjen, 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, 7.16.11–7.16.24 (2014).
21. Parts, L. et al. Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21, 1131–1138 (2011).
22. Wizneler, E. A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285, 901–906 (1999).
23. Leuenberger, P. et al. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* 355, eaal7825 (2017).
24. Wilkening, S. et al. An evaluation of high-throughput approaches to QTL mapping in *Saccharomyces cerevisiae*. *Genetics* 196, 853–865 (2014).
25. Kim, H. S., Huh, J., Riles, L., Reyes, A. & Fay, J. C. A noncomplementation screen for quantitative trait alleles in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 2, 753–760 (2012).
26. Sinha, H., Nicholson, B. P., Steinmetz, L. M. & McCusker, J. H. Complex genetic interactions in a quantitative trait locus. *PLoS Genet.* 2, e13 (2006).

Acknowledgements
The authors thank F. AlZaben, A. Flury, G. Geiselman, J. Hong, J. Kim, M. Maurer, and L. Oltrogge for technical assistance; D. Savage for his generosity with microscopy resources; and B. Blackman, S. Coradetti, A. Flamholz, V. Guacci, D. Koshland, C. Nelson, and A. Sasakura for discussions. We also thank J. Dusber (Department of Bioengineering, University of California, Berkeley) for the piggyBac cassette. This work was supported by R01 GM120430-A1 and by Community Sequencing Project 1460 to R.B.B. at the US Department of Energy (DOE) 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 US DOE under Contract No. DE-AC02-05CH11231.

Author contributions
R.B.B. and J.L.R. developed the project design; C.V.W., J.L.R., R.K.H., and J.N.C. performed experiments; C.V.W. and J.L.R. analyzed the data; J.M.S. contributed to the development of mutagenesis and sequencing methods; A.P.A. contributed mutagenesis and sequencing reagents; I.V.G. provided technical assistance with sequencing; and R.B.B. and C.V.W. wrote the manuscript with input from all authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0243-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.L.R. or R.B.B.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018
**Methods**

**Strains.** Strains used in this study are listed in Supplementary Table 2.

Homoygous 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 wildtype parents, each strain was rendered homozygous null for URA3 via homologous recombination with a HYGMX cassette, then sporulated; a mated spore from a dissected tetrad was grown 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* × *S. paradoxus* hybrid, we adapted methods for piggyBac transposon mutagenesis 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 1 we cloned the piggyBac transposase gene encoding transposon machinery into pTDH3 (provided by J. W. Carter, University of Notre Dame) via homologous arms (328 bp and 361 bp) from the piggyBac transposon. We first amplified KANMX from pU6G6 and each transposon arm from pE312, 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 2 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 gigaprepared using a column kit (Zymo Research) to generate ~11 mg of plasmid. To prepare for transformation, JR507 (the *S. cerevisiae* DBVPG1373 × *S. paradoxus* Z1 hybrid) was streaked from a ~80 °C freezer stock onto a yeast peptone dextrose (YPD; 1% yeast extract (BD Biosciences), 2% yeast peptone (BD Biosciences), 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 200 r.p.m. for 2 days at 26 °C. A single colony was inoculated into 100 ml YPD and shaken at 200 r.p.m. for 2 days at 26 °C. A single colony was inoculated into 100 ml YPD and shaken at 80 °C. We used this time point as time zero of our thermotolerance experiment.

**Tn-seq read-mapping and data analysis.** For analysis of data from the sequencing of transposon 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 transposon-containing read, we then identified the genomic location of the sequence immediately downstream of the transposon 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 corresponding to about 6.5 doublings in each case. Four cell pellets of 7 OD600 units each were homogenized and 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 3.0 fluorometer. Illumina transposon sequencing (Tn-seq) library construction was as described by Xu et al. (2015) with some modifications. Briefly, gDNA was fragmented and ligated with common adapters, and for each fragment deriving from a transposon insertion in the genome, the sequence containing a portion of the transposon and a portion of its genomic context (the transposon–genome junction) was amplified using one primer homologous to a region in the transposon and another primer homologous to a region in the adapter. See Supplementary Table 4 for the genomic primer (‘forward primer’), where NS represent random nucleotides, and the indexed adapter-specific primer (‘reverse primer’), where the six NS are a unique index used for multiplexing multiple libraries onto the same HiSeq sequencing lane. Amplification used Jumpstart polymerase (Sigma) and the following cycling protocol: 94 °C for 2 min, 94 °C for 30 s, 72 °C for 30 s at 30 cycles for 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 3.

**Transposon sequencing library construction.** To determine the abundance of transposon mutant hemizygote clones after selection, we first sequenced transposon 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 Research) prewarmed to 65 °C, and its concentration was quantified using a Qubit 3.0 fluorometer. Illumina transposon sequencing (Tn-seq) library construction was as described by Xu et al. (2015) with some modifications. Briefly, gDNA was fragmented and ligated with common adapters, and for each fragment deriving from a transposon insertion in the genome, the sequence containing a portion of the transposon and a portion of its genomic context (the transposon–genome junction) was amplified using one primer homologous to a region in the transposon and another primer homologous to a region in the adapter. See Supplementary Table 4 for the genomic primer (‘forward primer’), where NS represent random nucleotides, and the indexed adapter-specific primer (‘reverse primer’), where the six NS are a unique index used for multiplexing multiple libraries onto the same HiSeq sequencing lane. Amplification used Jumpstart polymerase (Sigma) and the following cycling protocol: 94 °C for 2 min, 94 °C for 30 s, 72 °C for 30 s at 30 cycles for 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 3.

**Thermotolerance phenotyping via selection of the hemizygote pool.** One aliquot of the pool of transposon mutant hemizygotes in the JR507 *S. cerevisiae* DBVPG1373 × *S. paradoxus* Z1 hybrid background was thawed and inoculated into 2 ml of YPD in a 250 ml flask and cultured at 27 °C. After 200 r.p.m. We used this time point as time zero of our thermotolerance experiment and treated four aliquots of 6.3 ml (7 OD600 units) as technical replicates for sequencing of transposon insertion positions (see below). Of the remaining culture, 9.19 ml was back-diluted to an OD600 of 0.02 in a total of 500 ml YPD in each of six 21 glass flasks for cultures that we call selections; three were grown at 28 °C and three at 39 °C (shaking at 200 r.p.m.) until an OD600 of 1.9–2.12 was reached, corresponding to about 6.5 doublings in each case. Four cell pellets of 7 OD600 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 and 39 °C of CEN6-initiated outside of the plasmid, and cultures were incubated with 200 r.p.m. shaking at 28 °C for 2 days.
selection, to yield a final abundance estimate for the insertion in this selection, \(<a_{\text{insert}}\>_{\text{total}}\). Likewise, for each insertion and selection experiment we calculated \(CV_{\text{insert, total}}\), the coefficient of variation of \(<a_{\text{insert}}\>_m\) 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 \(<a_{\text{insert}}\>_m\) value from the 39°C selection by the analogous quantity from the 28°C selection and took the log, 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 from 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 Fig. 5). 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), \(CV_{\text{insert, total}}\) ≥ 1.5 and \(a_{\text{insert}}\)_m ≥ 1.1, and genes for which this high-confidence insertion dataset contained at least five insertions in each species’ allele. This final dataset comprised 110,678 high-quality insertions (Supplementary Table 5), from the set of insertions falling into the S. cerevisiae alleles of each gene, and we tabulated single-replicate estimates of the abundances of hemizygotes harboring insertions in each species’ homolog; see columns G–L of Supplementary Table 6.

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 (see URLs). Reads were aligned using bowtie2 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 sequence11. SNPs were called using a pipeline of samtools12, bcftools, and bgzip and were filtered for a quality score of ≥20 and a combined depth of ≥5 and either \(D_{xy}\) (S. cerevisiae) or \(<\text{D}_{xy}\) (S. paradoxus). We then amended each reference genome with the respective filtered SNPs: we replaced the S288C allele with that of DBVPG1373 at each SNIP using bcftools’ consensus command with default options (42,983 bp total), and amendment of the CBS432 sequence was carried out analogously using Z1 alleles (15,126 bp total).

Sequence analysis. \(D_{xy}\) analysis. To evaluate whether sequence divergence from S. paradoxus at RH-seq hit loci was a shared feature of S. cerevisiae isolates, we used the \(D_{xy}\) statistic13, the average number of pairwise differences between S. cerevisiae strains and S. paradoxus, normalized for gene length, as follows. We downloaded S. cerevisiae genomic sequences from the following sources: YM978, UWOPS83-787, Y55, UWOPS05-2173, 273614N, YS9, BC187, YPS128, DBVPG6765, YM975, 13734, DBVPG1106, K11, SK1, 378604X, YM981, UWOPSS7-2421, DBVPG1373, NCCY3601, YP5606, Y12, UWOP505-227, and YS2 from the Yeast Resource Center (see URLs); Sigma1278, ZTW1, T7, and YM789 from the Saccharomyces Genome Database (see URLs); and RM11 from NCBI (accession PRNA13674). For each strain, we extracted the coding sequence of each gene in turn, and we downloaded the S. paradoxus reference sequence for each orthologous coding region from ref. 12. Sequences were aligned using MAFFT with default settings. Alignments that did not contain a start and stop codon, or those that contained gaps at greater than 40% of sites, were considered poor quality and were discarded. We tabulated \(D_{xy}\) for each gene. To evaluate whether the 8 RH-seq hit genes were enriched for elevated \(D_{xy}\), we first tabulated \(<D_{xy}>_{\text{total}}\), the mean value across the 8 RH-seq hit genes. We then sampled 8 random genes from the set of 3,416 genes tested by RH-seq; to account for biases associated with lower rates of divergence among essential genes, the resampled set contained 6 essential genes and 2 non-essential genes, mirroring the breakdown of essentiality among the RH-seq hits. Across this random sample we tabulated the mean \(D_{xy}\) of the 8 sampled genes. We repeated the resampling 5,000 times and used as an empirical P-value the proportion of resamples at which \(<D_{xy}>_{\text{sample}}\) ≤ \(<D_{xy}>_{\text{total}}\).

Phylogenetic analysis. We downloaded orthologous protein coding regions for the type strains of S. cerevisiae, S. paradoxus, and an outgroup, S. mikatae, from ref. 13. For each gene for which ortholog sequences were available in all three species, we aligned the sequences with PRANK14, utilizing the “-”-codon option for codon alignment. These alignments were used as input into the codeml module of PAML15, 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 8 RH-seq hit genes showed signatures of rapid evolution along the S. cerevisiae lineage, we used the resampling test detailed above.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Custom Python and R scripts used for RH-seq data analysis are available on GitHub (see URLs).

For strain construction, growth assays, microscopy, and locus effect size methods, see the Supplementary Note.

Data availability

RH-seq data have been deposited in the Sequence Read Archive (SRA) under accession SRP156210.

References

27. 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).

28. Wetmore, K. M. et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. mBio 6, e00306–e00315 (2015).

29. Skelly, D. A. et al. Integrative phenomenomics reveals insight into the structure of phenotypic diversity in budding yeast. Genome Res. 23, 1496–1504 (2013).

30. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

31. Liti, G. et al. Population genomics of domestic and wild yeasts. Nature 458, 337–341 (2009).

32. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

33. Nei, M. Molecular Evolutionary Genetics New York, (Columbia Univ. Press, 1987).

34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).

35. Loytynoja, A. & Goldman, N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics 11, 579 (2010).

36. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591 (2007).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size
   Describe how sample size was determined.

   Sample sizes were chosen based on technical limitations for a given experiment, e.g. the number of aliquots that could be managed in a transformation, or the number of glass tubes that could fit in a shaker.

2. Data exclusions
   Describe any data exclusions.

   No data were excluded from analysis.

3. Replication
   Describe whether the experimental findings were reliably reproduced.

   All growth phenotyping at high temperature was reproduced across at least two different days of experiments and for each transgenic, at least two genotypically identical strains were phenotyped.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.

   No randomization was required for this study. Samples that were separated into different treatment groups were randomly segregated.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   No blinding was required for the study.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   n/a Confirmed

   - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   - A statement indicating how many times each experiment was replicated
   - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   - The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
   - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   - Clearly defined error bars

   See the web collection on statistics for biologists for further resources and guidance.
Software

Describe the software used to analyze the data in this study.

Standard read-mapping and variant-calling software as noted in the methods section, plus custom software written by the authors in Python and R.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All genetically modified strains and plasmids created in this study are available upon request. Parent yeast strains are available from the National Collection of Yeast Cultures (NCYC). No other unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used in this study.

10. Eukaryotic cell lines

State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used in this study.

Describe the method of cell line authentication used.

No eukaryotic cell lines were used in this study.

Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used in this study.

If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used in this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

Provide details on animals and/or animal-derived materials used in the study.

No animals were used in this study.

Policy information about studies involving human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not involve human participants.