Journal of Visualized Experiments

Genetic mapping of thermotolerance differences between species of Saccharomyces yeast via genome-wide reciprocal-hemizygosity analysis

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE59972R1
Full Title:	Genetic mapping of thermotolerance differences between species of Saccharomyces yeast via genome-wide reciprocal-hemizygosity analysis
Keywords:	Genomics Yeast Evolution Reciprocal hemizygosity Genome-wide mapping Tn-seq
Corresponding Author:	Carly Weiss UC Berkeley Berkeley, California UNITED STATES
Corresponding Author's Institution:	UC Berkeley
Corresponding Author E-Mail:	rachelbrem@gmail.com
Order of Authors:	Carly Weiss
	Rachel Brem
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Berkeley, California, United States of America

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

May 3, 2019

To the editors:

This letter accompanies the submission of a revised manuscript, “Genetic mapping of thermotolerance differences between species of *Saccharomyces* yeast via genome-wide reciprocal-hemizygosity analysis,” by Carly Weiss and myself. The manuscript covers our methodology to map the mechanistic basis of trait variation between long-diverged species (Weiss *et al., Nature Genetics* 2018). Reviewers of our prevision submission were enthusiastic about the approach and about our detailed description of the workflow. In response to their suggestions, we have added to the protocol notes new data about expected performance of key steps; we also now expand on alternative strategies to several elements of the method, in protocol notes and the discussion.

We are confident that with these changes, our manuscript will now be suitable for publication in *JoVE*.

Thank you again for your expertise and hard work during the review process. I look forward to hearing from you.

Sincerely,

Rachel Brem, Ph.D.
Associate Professor
Buck Institute for Research on Aging

Adjunct Associate Professor
Department of Plant and Microbial Biology
University of California, Berkeley

Adjunct Associate Professor
Davis School of Gerontology
University of Southern California

8001 Redwood Boulevard
Novato, CA 94945
Phone: (415) 209-2093
Email: rbrem@buckinstitute.org

Manuscript Click here to access/download;Manuscript;59972_R1_051019.docx

TITLE:

Genetic Mapping of Thermotolerance Differences Between Species of *Saccharomyces* Yeast via Genome-Wide Reciprocal Hemizygosity Analysis

AUTHORS AND AFFILIATIONS:
Carly V. Weiss[1,2], Rachel B. Brem[1,3]

[1]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA
[2]Current address: Department of Biology, Stanford University, Stanford, CA, USA
[3]Buck Institute for Research on Aging, Novato, CA, USA

Corresponding author:
Rachel B. Brem (rbrem@buckinstitute.org)

Email addresses of co-author:
Carly V. Weiss (carly.weiss@berkeley.edu)

KEYWORDS:
Genomics, Yeast, Evolution, Reciprocal hemizygosity, Genome-wide mapping, Tn-seq

SUMMARY:
Reciprocal hemizygosity via sequencing (RH-seq) is a powerful new method to map the genetic basis of a trait difference between species. Pools of hemizygotes are generated by transposon mutagenesis and their fitness is tracked through competitive growth using high-throughout sequencing. Analysis of the resulting data pinpoints genes underlying the trait.

ABSTRACT:
A central goal of modern genetics is to understand how and why organisms in the wild differ in phenotype. To date, the field has advanced largely on the strength of linkage and association mapping methods, which trace the relationship between DNA sequence variants and phenotype across recombinant progeny from matings between individuals of a species. These approaches, although powerful, are not well suited to trait differences between reproductively isolated species. Here we describe a new method for genome-wide dissection of natural trait variation that can be readily applied to incompatible species. Our strategy, RH-seq, is a genome-wide implementation of the reciprocal hemizygote test. We harnessed it to identify the genes responsible for the striking high temperature growth of the yeast *Saccharomyces cerevisiae* relative to its sister species *S. paradoxus*. RH-seq utilizes transposon mutagenesis to create a pool of reciprocal hemizygotes, which are then tracked through a high-temperature competition via high-throughput sequencing. Our RH-seq workflow as laid out here provides a rigorous, unbiased way to dissect ancient, complex traits in the budding yeast clade, with the caveat that resource-intensive deep sequencing is needed to ensure genomic coverage for genetic mapping. As sequencing costs drop, this approach holds great promise for future use across eukaryotes.

INTRODUCTION:

Since the dawn of the field, it has been a prime goal in genetics to understand the mechanistic basis of variation across wild individuals. As we map loci underlying a trait of interest, the emergent genes can be of immediate use as targets for diagnostics and drugs, and can shed light on the principles of evolution. The industry standard toward this end is to test for a relationship between genotype and phenotype across a population via linkage or association[1]. Powerful as these approaches are, they have one key limitation—they rely on large panels of recombinant progeny from crosses between interfertile individuals. They are of no use in the study of species that cannot mate to form progeny in the first place. As such, the field has had little capacity for unbiased dissection of trait differences between reproductively isolated species[2].

In this work we report the technical underpinnings of a new method, RH-seq[3], for genome-scale surveys of the genetic basis of trait variation between species. This approach is a massively parallel version of the reciprocal hemizygote test[4,5], which was first conceived as a way to evaluate the phenotypic effects of allelic differences between two genetically distinct backgrounds at a particular locus (**Figure 1A**). In this scheme, the two divergent individuals are first mated to form a hybrid, half of whose genome comes from each of the respective parents. In this background, multiple strains are generated, each containing an interrupted or deleted copy of each parent's allele of the locus. These strains are hemizygous since they remain diploid everywhere in the genome except at the locus of interest, where they are considered haploid, and are referred to as reciprocal since each lacks only one parent's allele, with its remaining allele derived from the other parent. By comparing the phenotypes of these reciprocal hemizygote strains, one can conclude whether DNA sequence variants at the manipulated locus contribute to the trait of interest, since variants at the locus are the only genetic difference between the reciprocal hemizygote strains. In this way, it is possible to link genetic differences between species to a phenotypic difference between them in a well-controlled experimental setup. To date the applications of this test have been in a candidate-gene framework—that is, cases in which the hypothesis is already in hand that natural variation at a candidate locus might impact a trait.

In what follows, we lay out the protocol for a genome-scale reciprocal hemizygosity screen, using yeast as a model system. Our method creates a genomic complement of hemizygote mutants, by generating viable, sterile F1 hybrids between species and subjecting them to transposon mutagenesis. We pool the hemizygotes, measure their phenotypes in sequencing-based assays, and test for differences in frequency between clones of the pool bearing the two parents' alleles of a given gene. The result is a catalog of loci at which variants between species influence the trait of interest. We implement the RH-seq workflow to elucidate the genetic basis of thermotolerance differences between two budding yeast species, *Saccharomyces cerevisiae* and *S. paradoxus*, which diverged ~5 million years ago[6].

PROTOCOL:

1. Preparation of the piggyBac-containing plasmid for transformation

1.1. Streak out to single colonies the *E. coli* strain harboring plasmid pJR487 onto an LB +

89 carbenicillin agar plate. Incubate for 1 night at 37 °C or until single colonies appear.
90
91 NOTE: A description of how plasmid pJR487 was cloned can be found in our previous work[3].
92
93 1.2. Inoculate 1 L of LB + carbenicillin at 100 µg/mL with a single colony of *E. coli* containing
94 pJR487 in a 2 L glass flask. Grow overnight at 37 °C with shaking at 200 rpm until saturated (OD_{600}
95 ≥ 1.0).
96
97 1.3. Purify plasmid DNA from the culture using a large-scale plasmid prep kit as instructed in the
98 manufacturer's published protocol (see **Table of Materials** for details). Elute the DNA after a 10
99 minute incubation with 5 mL of elution buffer warmed to 37 °C.
100
101 1.4. Measure the quantity and quality of plasmid DNA with a spectrophotometer (see **Table of**
102 **Materials** for details).
103
104 1.5. Repeat steps 1.2 – 1.4 until a total of at least 11 mg plasmid DNA at an A_{260}:A_{280} ratio of at
105 least 1.8 are isolated. This may take a few preps, depending on efficiency.
106
107 1.6. Mix all plasmid preps together into a single tube and bring the total volume up to 20 mL with
108 elution buffer or water. Measure the final quantity and quality again with a spectrophotometer.
109 The concentration of plasmid should be at least 538 ng/µL in this final 20 mL volume. If the
110 concentration is higher than 538 ng/µL, dilute the plasmid with elution buffer or water to 538
111 ng/µL. Plasmid can be stored at 4 °C up to a few weeks until use.
112
113 **2. Creating a pool of untargeted genome-wide reciprocal hemizygotes**
114
115 2.1. Preparation of hybrid yeast cells for transformation
116
117 2.1.1. Streak out JR507 from a -80 °C freezer stock strain to single colonies on a YPD agar plate.
118 Incubate at 26 °C for 2 days or until colonies appear.
119
120 NOTE: JR507 is a hybrid strain made through single-cell mating of haploid spores of S. cerevisiae
121 DBVPG1373 and S. paradoxus Z1 (using a tetrad-dissection microscope)[3].
122
123 2.1.2. Inoculate 100 mL of liquid YPD in a 250 mL glass flask with a single colony of JR507 and
124 shake at 28 °C, 200 rpm for 24 hours, or until stationary phase is reached.
125
126 2.1.3. The next day, measure the optical density at 600 nm (OD_{600}) of the overnight culture.
127 Create a new culture by back-diluting some of the overnight culture with fresh liquid YPD into a
128 new 1 L glass flask to an OD_{600} of 0.2 and a volume of 500 mL.
129
130 NOTE: Example calculation of a back-dilution if the overnight culture has an OD_{600} of 5.0, where
131 C is optical density and V is volume:
132

$$C_1 * V_1 = C_2 * V_2$$
$$(5.0) * V_1 = (0.2) * (500 \text{ mL})$$
$$V_1 = [\ (0.2) * (500 \text{ mL})\ ] / (5.0)$$
$$V_1 = 20 \text{ mL}$$

Thus, 20 mL of saturated overnight culture would be added to 480 mL of liquid YPD to make a total of 500 mL of culture at an OD_{600} of 0.2.

2.1.4. Repeat step 2.1.3 three more times to make a total of four 500 mL cultures at an OD_{600} of 0.2 in four 1 L glass flasks, using the same overnight culture for all four new cultures. Incubate them all at 28 °C for 6 hours (2-3 generations) shaking at 200 rpm.

2.1.5. Combine two of the 500 mL cultures to create a 1 L culture. Combine the remaining two 500 mL cultures to create another 1 L culture. At this point, there are two 1 L cultures. Each of these 1 L cultures will be subject to transformation with pJR487 in the following steps.

2.2. Transformation of pJR487 into hybrid yeast cells

2.2.1. Split each of the 1 L cultures into twenty 50 mL aliquots in 20 plastic conical tubes for a total of 40 tubes. Set aside 20 tubes and perform the following steps on 20 tubes at a time.

2.2.2. Centrifuge each of the twenty tubes for 3 min at 1,000 x *g* to pellet the yeast cells. Discard the supernatant.

2.2.3. Resuspend each pellet with 25 mL of sterile H_2O by vortexing. Centrifuge for 3 min at 1,000 x *g*. Discard the supernatant.

2.2.4. Resuspend each pellet with 5 mL of 1x TE, 0.1 M LiOAc buffer by vortexing. Centrifuge for 3 min at 1,000 x *g*. Discard the supernatant.

2.2.5. Repeat step 2.2.4. While the cells are centrifuging, prepare at least 120 mL of a solution of 39.52% polyethylene glycol, 0.12 M LiOAc and 1.2x Tris-EDTA buffer (12 mM Tris-HCl and 1.2 mM EDTA). Store on ice.

2.2.6. To prepare the plasmid DNA for transformation, first boil 4 mL of salmon sperm DNA at 100 °C for 5 min and immediately cool it on ice for 5 min. Then, mix 20 mL of pJR487 (obtained in section 1) at a concentration of 538 ng/µL with the 4 mL of cooled salmon sperm DNA for a total volume of 24 mL. Keep on ice until use.

2.2.7. Add 600 µL of plasmid DNA mixed with salmon sperm DNA on top of each cell pellet. Do not resuspend yet.

2.2.8. Add 3 mL of PEG-LiOAc-TE solution made in step 2.2.5 to each pellet. Resuspend the pellet by pipetting up and down and vortexing.

2.2.9. Incubate each tube for 10 min at room temperature.

2.2.10. Heat shock each tube for 26 min in a water bath set to 39 °C.

NOTE: Every few minutes, invert each tube to prevent the cells from settling on the bottom of the tube.

2.2.11. Centrifuge each tube for 3 min at 1,000 x *g*. Discard the supernatant and resuspend each pellet in 10 mL of YPD by vortexing. Combine all twenty tubes into a new glass flask. The total volume of cells should be ~200 mL.

2.2.12. Transfer 66.6 mL of cells to a new 1 L glass flask and bring up to a volume of 500 mL with liquid YPD. Repeat two more times to use the entire 200 mL of transformed cells. Measure the OD_{600} of each new 500 mL culture (expect an OD_{600} of ~0.35-4).

2.2.13. Shake all three flasks at 28 °C for 2 hours to recover (<1 generation) at 200 rpm.

2.2.14. Add 0.5 mL of 300 mg/mL G418 to each of the three flasks, to a final concentration of 300 µg/mL G418 and put back to shake at 28 °C, 200 rpm.

NOTE: Prior to this step, the transformed hybrid cells have been recovering from transformation. Upon the addition of G418, presence of the plasmid pJR487 is selected for. Any cells that did not take up the plasmid during transformation will begin to die.

2.2.15. Repeat steps 2.2.2 – 2.2.14 with the remaining 20 conical tubes of cells. At this point there should be six 1 L glass flasks, each with 500 mL of cells with G418 added.

2.2.16. Incubate all six flasks of cells at 28 °C, shaking at 200 rpm, for approximately 2 days or until an OD_{600} of ~2.3 is reached in each flask. Combine all six flasks together to create a single culture.

NOTE: Although all of the cells in this culture will not be used in downstream steps, the goal of using such large volumes has been to create as many unique transformation events as possible and normalize any biases across a single transformation by pooling them all together.

2.2.17. Use the culture created in 2.2.16 to inoculate two new 1 L flasks with 500 mL of YPD + G418 (300 µg/mL) to an OD_{600} of 0.2. There will be leftover culture that can be discarded.

2.2.18. Incubate both 1 L flasks at 28 °C overnight, with shaking at 200 rpm, until each reaches an OD_{600} of ~2.2 (~3.5 generations). Combine both cultures into a single culture and measure the OD_{600} of the combined culture again.

NOTE: At this point, the culture should be almost entirely comprised of cells harboring plasmid

221 pJR487. In part of the population of cells, the PiggyBac transposon will have been transposed
222 from the plasmid into the genome by the transposase expressed off the plasmid. However,
223 continued expression of the transposase can lead to transposition during the course of a
224 selection, which would obscure the relationship between genotype and phenotype. The goal of
225 the next several steps is to perform a counterselection against the presence of the plasmid, to
226 ensure there is no more expression of the transposase. The resulting pool is a mix of cells with or
227 without the transposon integrated into the genome, but only cells containing the transposon are
228 detected during the subsequent mapping steps. The time in the transformation during which
229 transposase is expressed, before the plasmid encoding is lost, may govern the chance that a given
230 clone after mutagenesis harbors more than one transposon insertion. The frequency of these,
231 which manifest as "secondary" mutations in analysis of any one gene at a time, can be estimated
232 by arraying a defined number of colonies after mutagenesis, then combining their DNA and
233 sequence-confirming the number of independent insertion positions in the pool.
234
235 2.2.19. Centrifuge 25 mL of this culture for 3 min at 1,000 x *g*. Calculate the number of total OD_{600}
236 units of cells that are in the 25 mL (see example calculation below). Discard the supernatant and
237 resuspend in enough H_2O to create a cell suspension of 1.85 OD_{600}/mL by vortexing.
238
239 NOTE: Example calculation for resuspension of cells in water if OD_{600} of combined culture was
240 2.2:
241
242 2.2 OD_{600} units of cells * 25 mL = 55 OD_{600} units of cells
243 55 D_{600} units of cells / *x* mL H_2O = 1.85 OD_{600} units of cells / mL
244 *x* mL H_2O = 55 OD_{600} units of cells 1.85 OD_{600} units of cells / mL
245 *x* = 29.7 mL
246
247 So, after spinning 25 mL of cell culture and discarding the supernatant, add enough H_2O to the
248 cells to bring the total volume of cells and water up to ~29.7 mL (since the cell pellet will also
249 have a volume, add less than 29.7 mL of H_2O).
250
251 2.2.20. Using glass beads, plate 1 mL of resuspended cells in water onto each of 12 large square
252 complete synthetic agar plates with 5-FOA. Incubate each plate at 28 °C for 1-2 days or until a
253 lawn forms on the plate.
254
255 2.2.21. Using small sterile squeegees, scrape the cells off of each of 6 plates and into a tube with
256 35 mL of sterile water. Repeat with the other 6 plates for a total of two tubes of cells and water.
257 Combine all cell suspensions in a single tube. Measure the OD_{600} of this suspension, using water
258 as a blank. Bring the OD_{600}/mL concentration of cells to 44.4 OD_{600} units/mL with water. In our
259 experience, transposition efficiency (the proportion of KAN+ cells that are URA-) is on average
260 50%.
261
262 2.2.22. Determine the number of -80 °C freezer stocks of cells to store. Each aliquot can be used
263 in the future for a single experiment.
264

265 NOTE: Given how time consuming the generation of the pool is, store multiple vials in case of
266 accidental misuse or for performing replicate experiments. 20-30 stocks are a reasonable
267 number.
268
269 2.2.23. Each freezer stock will contain 40 OD_{600} units of cells in 1 mL of 10% DMSO. Add 900 µL
270 of cells to 100 µL of DMSO. Repeat for the total number of freezer stocks created. Store each at
271 -80 °C for future use.
272
273 **3. Selection of reciprocal hemizygotes in a pooled format**
274
275 3.1. Thaw from the -80 °C freezer a single aliquot of pooled reciprocal hemizygotes from section
276 2 at room temperature.
277
278 NOTE: Do not let the aliquot sit for long at room temperature once it thaws, use it immediately.
279
280 3.2. Use the entire 1 mL aliquot to inoculate 150 mL of liquid YPD in a 250 mL glass flask. Measure
281 the OD_{600} of this culture, and then incubate at 28 °C, shaking at 200 rpm, for ~7 hours, or until
282 the culture has gone through 2-3 population doublings. At this point, the culture is ready to be
283 used to inoculate cultures undergoing selection.
284
285 NOTE: Example calculation: If the OD_{600} of the original flask measures 0.25, incubate the culture
286 until it reaches an OD_{600} of at least 1.0. If any sample points are desired at "Time-zero" (T-0), as
287 a way to investigate the hemizygote population before selection, cell pellets can be taken now
288 by centrifuging 5-10 mL of culture per pellet at 1,000 x *g* for 3 min, discarding the supernatant
289 and freezing at -80 °C.
290
291 3.3. Use the grown hemizygote pool to inoculate cultures for selection in a suitable replicate
292 scheme, at both high temperature (39 °C) and permissive temperature (28 °C). At a minimum,
293 set up three biological replicate selection cultures at each temperature, for a total of six selection
294 cultures.
295
296 3.3.1. Create each selection culture with 500 mL total in a 2 L glass flask with liquid YPD and
297 inoculate to an OD_{600} of 0.02. Shake each selection culture at 100 rpm at either 28 °C or 39 °C
298 until 6-7 population doublings have occurred (corresponding to an OD_{600} of ~1.28-2.56). Try to
299 match as closely as possible the final OD_{600} of all selection cultures.
300
301 NOTE: Selection cultures at 28 °C will grow faster than selection cultures at 39 °C. Consequently,
302 selection cultures at 39 °C will spend a longer period of time in the incubator. Proceed with the
303 following steps with each flask as it becomes ready, regardless of the total number of hours spent
304 in the incubator. In our experience, cultures at 28 °C or 39 °C took ~12 or ~18 hours, respectively,
305 to reach an OD of ~2.0. Long selections could have the advantage of amplifying small fitness
306 effects, but also permit de novo background mutations to arise, which would introduce noise into
307 the final distribution of fitnesses across transposon mutants in any one gene/allele. As such it is
308 important to limit selection time in an RH-seq experiment.

3.4. Harvest cell pellets from each selection culture. Calculate the volume required to obtain 7 OD_{600} units of cells and centrifuge at 1,000 x *g* for 3 min at least four pellets of this volume from each selection culture as technical replicates for library preparation and sequencing (see sections 4 and 5, below). Discard the supernatant and store at -80 °C.

NOTE: Example if a selection flask has a final OD_{600} of 2.0:

$$2.0\ OD_{600} \text{ units of cells / mL} = 7\ OD_{600} \text{ units of cells / } x \text{ mL}$$

$$x \text{ mL} = 3.5 \text{ mL of culture to centrifuge for a pellet of } 7\ OD_{600} \text{ units of cells.}$$

4. Tn-seq library construction and Illumina sequencing to determine abundance of transposon mutant hemizygotes

4.1. Thaw on ice each cell pellet from section 3 that is going to be sequenced.

4.2. Isolate total genomic DNA (gDNA) from each cell pellet using a yeast gDNA purification kit following the manufacturer's instructions. Resuspend the DNA in 50 µL of elution buffer warmed to 65 °C.

4.3. Quantify the quantity of gDNA from each pellet using a fluorimeter. The minimum total quantity of gDNA required for each cell pellet to create a next-generation sequencing (NGS) library for Tn-seq using the following procedure is 1 µg.

NOTE: Less than 1 µg of gDNA can be used to create a library, but the final quantity and quality of the library will suffer.

4.4. Follow an established protocol for creating Tn-seq libraries[7]. Note the following relevant information that is unique to this protocol:

4.4.1. After gDNA shearing, end repair and adapter ligation, amplify the gDNA containing the transposon via PCR. For that PCR, use the following forward and reverse primer, which are specific for the PiggyBac transposon and NGS adapters, respectively:
Forward (N – random nucleotide)
5' ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG
CTCTTCCGATCTNNNNNNAGCAATATTTCAAGAATGCATGCGTCAAT 3'
Reverse (the stretch of Ns represents a unique 6-bp index used for multiplexing. See below for further information on indices)
5' CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAG
ACGTGTGCTCTTCCGATCT 3'

4.4.2. Use the included cleanup steps with size-selective beads to minimize the proportion of cloned fragments in the final library that would be too short to include mappable genomic sequence.

NOTE: Having followed the minimum replicate requirements up until now for selection cultures, there will be 24 individual gDNA samples for sequencing. Given the current cost for sequencing, it is unlikely that each sample will be run on its own. To combine samples on the same lane, create multiple reverse primers, each with a unique 6-base pair index. Samples with differing indexes can be combined into the same sequencing lane and separated computationally afterwards.

4.5. Sequence single-end 150 bp reads from each library using NGS technologies across eight lanes.

NOTE: The amount of sequencing reads required depends heavily on the quality of the libraries prepared in the previous step (i.e. the proportion of DNA in the library actually containing transposon DNA, representing DNA coming from reciprocal hemizygotes). There are two main factors contributing to this. First, since cells without an integrated transposon are not counterselected against during pool creation, each culture will be a mix of cells with and without the transposon. Secondly, even within the genomes of transposon-containing reciprocal hemizygotes, most of the genome is not transposon containing sequence, and this gDNA will unavoidably be part of the library preparation. The goal of the final PCR amplification of transposon-containing DNA is to increase the ratio of transposon-containing DNA to these two sources of background gDNA. The more efficient this amplification is, the higher proportion of reads will be able to be used in downstream analysis. The lower quality the libraries are, the more sequencing will need to be done, since an increasing proportion of reads will not contain transposon DNA and will not be useful. Given the above constraints, eight lanes of sequencing were capable of tracking reciprocal hemizygote abundances to a reasonable degree. More sequencing would allow a deeper analysis.

5. Mapping the locations of transposon insertions and RH-seq analysis

NOTE: The following data analysis was accomplished with custom Python scripts (found online at https://github.com/weiss19/rh-seq), but could be redone using other scripting languages. Below, the major steps in the process are outlined. Perform the following steps on each individual replicate read file unless it is noted to combine them.

5.1. Trim adapter sequences off of reads and separate out each replicate's reads according to index.

5.2. Find reads containing transposon-genome junctions. To accomplish this, search within each read for the last 20 base pairs of the transposon, CAGACTATCTTTCTAGGGTTAA. Discard all reads not containing this sequence.

NOTE: In our experience, the proportion of reads mapping to the end of the transposon is 83-95%.

397 5.3. Trim the remaining, transposon-containing reads to contain only the sequence downstream
398 of the 3' end of the transposon. By mapping this sequence to the yeast genome, determine the
399 genomic context of the transposon insertion for each read (step 5.4 below).
400
401 5.4. Use BLAT or an equivalent mapping tool to map the sequence downstream of the transposon
402 to the *S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1 hybrid genome (Script name:
403 map_and_pool_BLAT.py).
404
405 5.4.1. Discard any reads for which there are fewer than 50 base pairs of usable sequence
406 downstream of 3' end of the transposon. Short sequences are difficult to map uniquely.
407
408 5.4.2. If using BLAT, use the following parameters: identity = 95, tile size = 12.
409
410 5.4.3. Create a basic hybrid genome to use for mapping by concatenating the latest versions of
411 reference genomes of S. cerevisiae S288c and S. paradoxus CBS432.
412
413 NOTE: A basic annotation file describing the genomic boundaries of individual genes across the
414 hybrid genome can be found at the Github repository listed above (Filename:
415 YS2+CBS432+plasmid_clean). Only use reads which map to a single location in the hybrid genome
416 (i.e. are unique to either S. cerevisiae or S. paradoxus). A uniform frequency of insertion events
417 across the genome is expected; the distribution of insertion positions across the genomes is
418 reported elsewhere[3].
419
420 5.5. Tally the total number of reads mapping to each unique transposon insertion location, which
421 we infer all originated from cells of a single transposon insertion mutant clone. The sum of all
422 such values from a single library is referred to as the total number of mapped reads for that
423 library.
424
425 5.6. In cases where there are multiple insertions mapping within 3 base pairs of one another,
426 combine them all to a single insertion point, assigning all the reads to the single location with the
427 highest read count. This value, n_{insert}, represents the abundance of that insertion clone in the cell
428 pellet from which gDNA was sequenced. At this point, there will be lists of n_{insert}, each the
429 abundance of a unique mapped transposon insertion, one list for every cell pellet sequenced.
430
431 NOTE: The PiggyBac transposon inserts at TTAA sequences in the genome, a 4 base pair sequence.
432 Thus, we infer that insertions mapping within 3 base pairs of each other must have originated
433 from the same TTAA site.
434
435 5.7. Since there will be a slightly different number of total reads coming from each sequenced
436 library, normalize the n_{insert} values across all files if they are to be compared. Do so by tabulating
437 the total number of mapped reads from each individual library, n_{pellet}, and take the average of all
438 n_{pellet} across all libraries, $\langle n_{pellet} \rangle$. Multiply each n_{insert} in an individual library's data by the ratio of
439 $\langle n_{pellet} \rangle$ / n_{pellet} to calculate a_{insert}, the normalized abundance of a given transposon insertion
440 clone.

$$a_{insert} = n_{insert} * (\langle n_{pellet} \rangle / n_{pellet})$$

Alternatively, library size can be estimated using available tools like DESeq2[8] (Script name: total_reads_and_normalize.py).

5.8. Tabulate the set of all insertions mapped across all libraries. For insertions found in some libraries but not in others, set $a_{insert} = 1$ for downstream calculations.

5.9. Filter the reads to find those insertions that fall within genes according to the annotation file (Script name: remove_NC_and_plasmid_inserts.py).

5.10. For each unique insertion, calculate the average abundance across technical replicates of each selection (each culture at either 28 °C or 39 °C), $\langle a_{insert} \rangle_{technical}$ (Script name: combine_tech_reps_V2.py).

5.11. For each unique insertion, calculate the average abundance across biological replicates of each temperature, $\langle a_{insert} \rangle_{total}$, by taking the mean of all $\langle a_{insert} \rangle_{technical}$ at each temperature. At the same time, calculate the coefficient of variation for each insertion, $CV_{insert,total}$ across the $\langle a_{insert} \rangle_{technical}$ (Script name: combine_bio_reps.py).

NOTE: At this point, for each temperature, 28 °C and 39 °C, there is a list of unique transposon insertions, their average abundance and the coefficient of variation between biological replicates for each. These data for our experiment are reported elsewhere[3].

5.12. Filter the list of all insertions for those that have, at either 28 °C or 39 °C, $\langle a_{insert} \rangle_{total} > 1.1$, and $CV_{insert,total} \leq 1.5$ (Script name: filter_inserts.py).

5.13. For each unique insertion, calculate the $\log_2(\langle a_{insert} \rangle_{total,28\,°C} / \langle a_{insert} \rangle_{total,39\,°C})$. This value represents the "thermotolerance" of a given transposon insertion mutant clone (Script name: fitness_ratios.py).

5.14. Sort all of the unique insertions by gene and by allele (*S. cerevisiae* or *S. paradoxus*), and tabulate the number of insertions in each allele. Filter genes so that only genes that have at least 5 insertions in each allele are analyzed (Script name: organize_and_filter_genes.py).

NOTE: Multiple unique insertions across each allele allow for a more accurate measure of that reciprocal hemizygote's thermotolerance. Lowering the number of insertions required per allele is possible but will compromise the accuracy of this measure and increase the multiple testing burden by allowing more genes to be tested. Additionally, filtering out genes with too few insertions per allele will help reduce the impact on test results of any individual hemizygote clone harboring a secondary site mutation that confers a very disparate phenotype.

5.15. For each gene remaining in the data set after the above filtering, compare the

thermotolerances ($\log_2$ ratios) of all the insertions in the *S. cerevisiae* allele to those in the *S. paradoxus* allele using a Mann-Whitney U test. Alternatively, a regression model could be implemented, adapted from DESeq2[8] (Script name: mann_whitney_u.py).

5.16. Correct *p*-values for multiple testing using the Benjamini-Hochberg method.

5.17. Genes with significant *p*-values (say, ≤ 0.01) are candidates for genes important for differences in thermotolerance between the two species.

REPRESENTATIVE RESULTS:

We mated *S. cerevisiae* and *S. paradoxus* to form a sterile hybrid, which we subjected to transposon mutagenesis. Each mutagenized clone was a hemizygote, a diploid hybrid in which one allele of one gene is disrupted (**Figure 1A, Figure 2**). We competed the hemizygotes against one another by growth at 39 °C and, in a separate experiment as a control, at 28 °C (**Figure 1B**), and we isolated DNA from each culture. To report the fitness of each hemizygote we quantified abundance via bulk sequencing, using a protocol in which DNA was fragmented and ligated to adapters, followed by amplification of transposon insertion positions (**Figure 1C**). If the primers for this amplification are distinct from, and less efficient than, those provided in the protocol, background reads will predominate in the sequencing data, leading to fewer usable reads and eroding the accuracy of fitness estimates. Similar quality issues may result from low DNA input into the sequencing library preparation.

With results in hand from our sequencing, for a given gene we compared hemizygote abundances at the two temperatures between two classes of hemizygotes: clones where only the *S. cerevisiae* allele was wild-type and functional, and clones relying only on the *S. paradoxus* allele (**Figure 1D**). In analysis at this stage, if the computational post-processing strategy of the protocol is not followed and genes with relatively few transposon mutants in the pool are included in the analysis, statistical power will drop and no significant gene calls will result. In our implementation, we detected strong signal at eight housekeeping genes (**Figure 3**). In each case, transposon insertions in the *S. cerevisiae* allele in the hybrid compromised growth at high temperature (**Figure 3**). These loci represented candidate determinants of the thermotolerance trait that distinguishes *S. cerevisiae* from *S. paradoxus*. In separate experiments reported elsewhere, we validated the impact of allelic variation at each site using standard transgenesis methods beyond the scope of the current protocol[3].

FIGURE AND TABLE LEGENDS:

Figure 1. Schematic of the RH-seq workflow. A. *S. cerevisiae* and *S. paradoxus* (blue and yellow respectively), are mated to form a hybrid (green) that contains a single copy of each of the parents' genomes. At a given locus in the hybrid, a transposon insertion (black box) in each species' allele in turn creates a hemizygote, which is diploid at the rest of the genome except for the locus of interest. Comparing phenotypes across hemizygotes reveals the phenotypic effects of allelic variation at the manipulated locus. **B.** Across many clones hemizygous at a given gene (*YFG*), some reach higher abundance than others in competitive culture, as quantified by

529 sequencing. **C.** DNA from a hemizygote pool is sheared and ligated to adapters (red). For a given
530 clone, the junction between the transposon (tn, black) and the genome (blue) is amplified with a
531 transposon-specific primer (black arrowhead) and an adapter-specific primer (red arrowhead).
532 Sequencing read counts from the amplicon report the fitness of the clone in the population.
533 **D.** For an RH-seq gene hit, tabulating the proportion of hemizygote clones (*y*-axis) exhibiting a
534 given fitness after competition at high temperature (*x*-axis) reveals a striking difference between
535 two genotypic classes: those with a transposon insertion in the *S. cerevisiae* allele (with the *S.*
536 *paradoxus* allele remaining; yellow) and those with the *S. paradoxus* allele disrupted (and *S.*
537 *cerevisiae* allele remaining; blue).
538
539 **Figure 2. Selection scheme for generating a pool of genome-wide reciprocal hemizygotes with**
540 **the PiggyBac plasmid-borne transposon.** The PiggyBac plasmid (pJR487) is transformed into a
541 *URA3*$^{-/-}$ clone of the diploid hybrid *S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1 (JR507). The
542 presence of the plasmid or transposon is selected for via growth in G418, which selects for the
543 presence of the KanMX cassette; survivors are cells which have taken up the PiggyBac plasmid
544 and/or harbor an integrated transposon. Cells without the latter are selected against via growth
545 in 5-FOA, which is toxic in the presence of the *URA3* cassette. Since the untransformed hybrid is
546 *URA3*$^{-/-}$, the only cells that will die in this step are those still containing the PiggyBac plasmid,
547 which contains a *URA3* cassette. What remains is a pool of hybrid mutant cells containing the
548 transposon integrated into the genome.
549
550 **Figure 3. Top hits mapped by RH-seq.** Each panel reports RH-seq data for the indicated gene
551 from RH-seq. The *x*-axis reports the $\log_2$ of abundance of a transposon mutant clone after
552 selection at 39 °C, relative to the analogous quantity at 28 °C. The *y*-axis reports the proportion
553 of all clones bearing insertions in the indicated allele that exhibited the abundance ratio on the *x*,
554 as a kernel density estimate. Underlying read and count data for insertions are reported
555 elsewhere[3].
556
557 **DISCUSSION:**
558 The advantages of RH-seq over previous statistical-genetic methods are several-fold. In contrast
559 to linkage and association analysis, RH-seq affords single-gene mapping resolution; as such, it will
560 likely be of significant utility even in studies of trait variation across individuals of a given species,
561 as well as interspecific differences. Also, previous attempts at genome-wide reciprocal
562 hemizygosity analysis used collections of gene deletion mutants, some of which harbor
563 secondary mutations that can lead to false positive results[9,10]. The RH-seq strategy sidesteps this
564 issue by generating and phenotyping many hemizygote mutants in each gene in turn, such that
565 the background of any individual mutant clone contributes only marginally to the final result. In
566 principle, RH-seq also affords the study of noncoding loci, although in the current work we
567 focused exclusively on genes.
568
569 There are a few quirks to RH-seq, some biological and some technical, that a successful
570 practitioner will deal with up front to maximize the utility of the approach and accelerate the
571 path to best results. Biologically, RH-seq only makes sense as a technique if the two target species
572 can be mated to form a stable, viable hybrid that can be genetically manipulated. Thus we cannot

envision applying RH-seq to species so divergent that they fail to fuse into a karyotypically stable diploid. On the other hand, if the two parents of the diploid hybrid are too similar at the DNA level, most reads from the transposon insertion sequencing cannot be mapped allele-specifically to just one of the two parent genomes and will be unusable; thus, a given RH-seq experiment will be most successful when the parents have high-quality reference genomes available and hit a "sweet spot" of sequence divergence. As a separate point of consideration, given an RH-seq project formulated to dissect the genetic basis of a trait difference between the parent species, results are likely to be much more interpretable when, for the trait of interest, the biology of the hybrid serves as a reasonable representative of that of the parents. Extreme phenotypes unique to the hybrid (heterosis) could influence or obscure the effects of genes of interest underlying the phenotype as it differs between the parents. Any genes mapped through reciprocal hemizygosity analysis must be validated by independent allele-swap experiments in the genetic backgrounds of the purebred parent species.

As for technical issues in an RH-seq experiment, our experience has highlighted several potential sources of noise and provided workable solutions. Noise manifests as disagreement among the sequencing-based estimates of fitness of the hemizygotes harboring transposon insertions in a given allele of a given gene. This can derive from differing secondary mutations in the backgrounds of transposon mutants (see below); variability in the efficiency of the PCR amplifying different insertion sites; low representation of a given mutant in the bulk pool, leading to low sequencing coverage which weakens precision; and differences in position of the transposon insertion within the gene (e.g., transposons inserting at a 3' gene end may have minimal phenotypic effect). For all these reasons, we consider it critical to generate very large transposon mutant pools and, in the final analysis, to exclude from testing any gene without a reasonable number of mutants in each of the two alleles. We note that, although we have not implemented it here, a barcoded transposon system[7] could further help resolve issues of PCR bias and cut down on the cost and labor of an RH-seq experiment.

In conclusion, we have established a straightforward workflow for RH-seq, and have specified caveats of the approach. We find that the latter does not significantly compromise the utility of RH-seq; we consider that it holds great promise for high-resolution, genome-scale dissection of the phenotypic consequences of genetic variation, including differences between species that have been reproductively isolated for millions of years.

ACKNOWLEDGMENTS:
We 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. Sasikumar for discussions; we also thank J. Dueber (Department of Bioengineering, UC Berkeley) for the PiggyBac plasmid. 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.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Flint, J., Mott, R. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews Genetics*. **2**, 437-45 (2001).
2. Allen Orr, H. The genetics of species differences. *Trends in Ecology and Evolution*. **16**, 343-350 (2001).
3. Weiss, C.V. et al. Genetic dissection of interspecific differences in yeast thermotolerance. *Nature Genetics*. **50**, 1501-1504 (2018).
4. Stern, D.L. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *Trends in Genetics*. **30**, 547-54 (2014).
5. Steinmetz, L.M. et al. Dissecting the architecture of a quantitative trait locus in yeast. *Nature*. **416**, 326-30 (2002).
6. 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 (2011).
7. Wetmore, K.M. et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio*. **6**, e00306-15 (2015).
8. Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. **15**, 550 (2014).
9. Wilkening, S. et al. An evaluation of high-throughput approaches to QTL mapping in Saccharomyces cerevisiae. *Genetics*. **196**, 853-65 (2014).
10. 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**, 753-60 (2012).

Figure 1

Click here to access/download;Figure;Figure_1.png

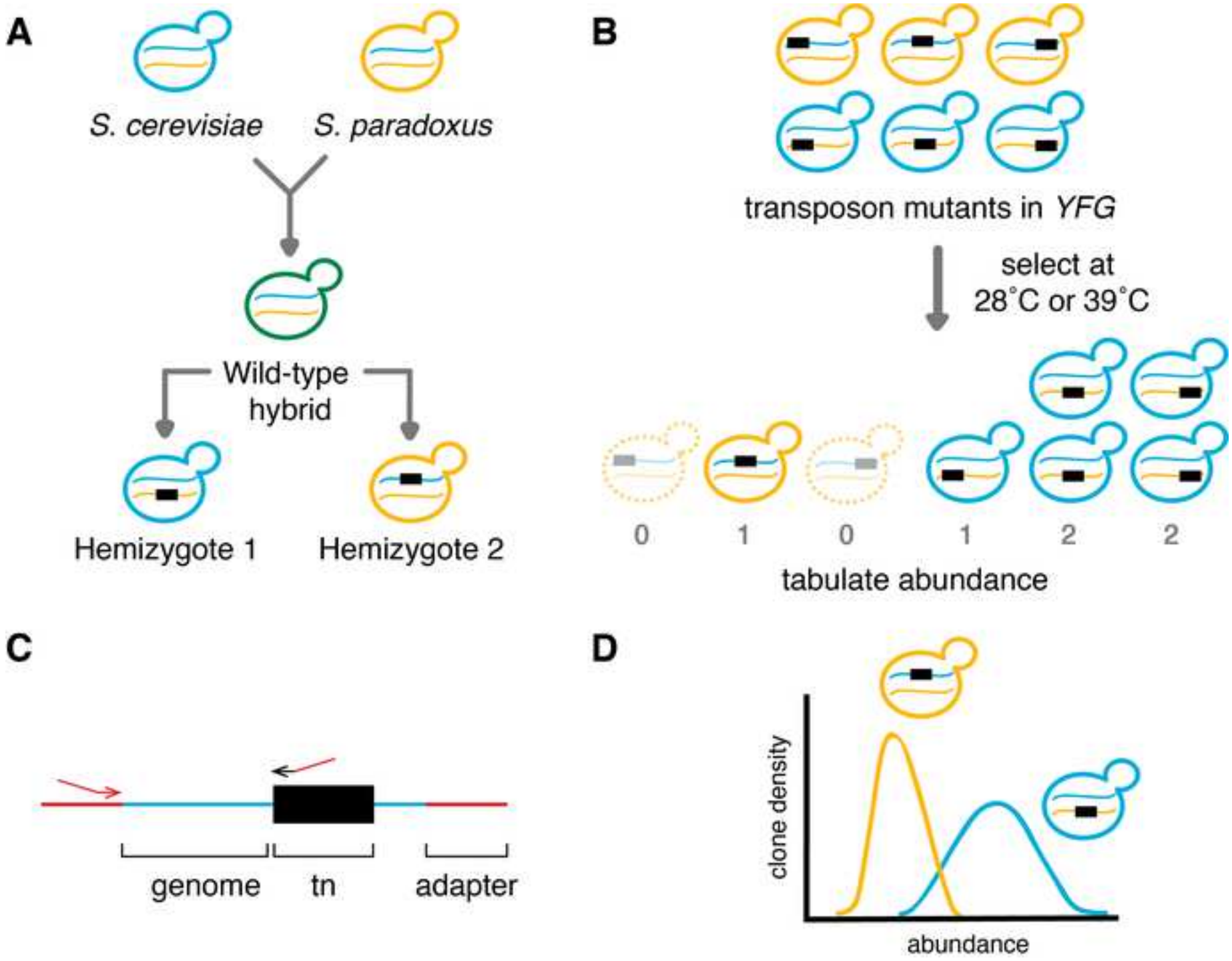
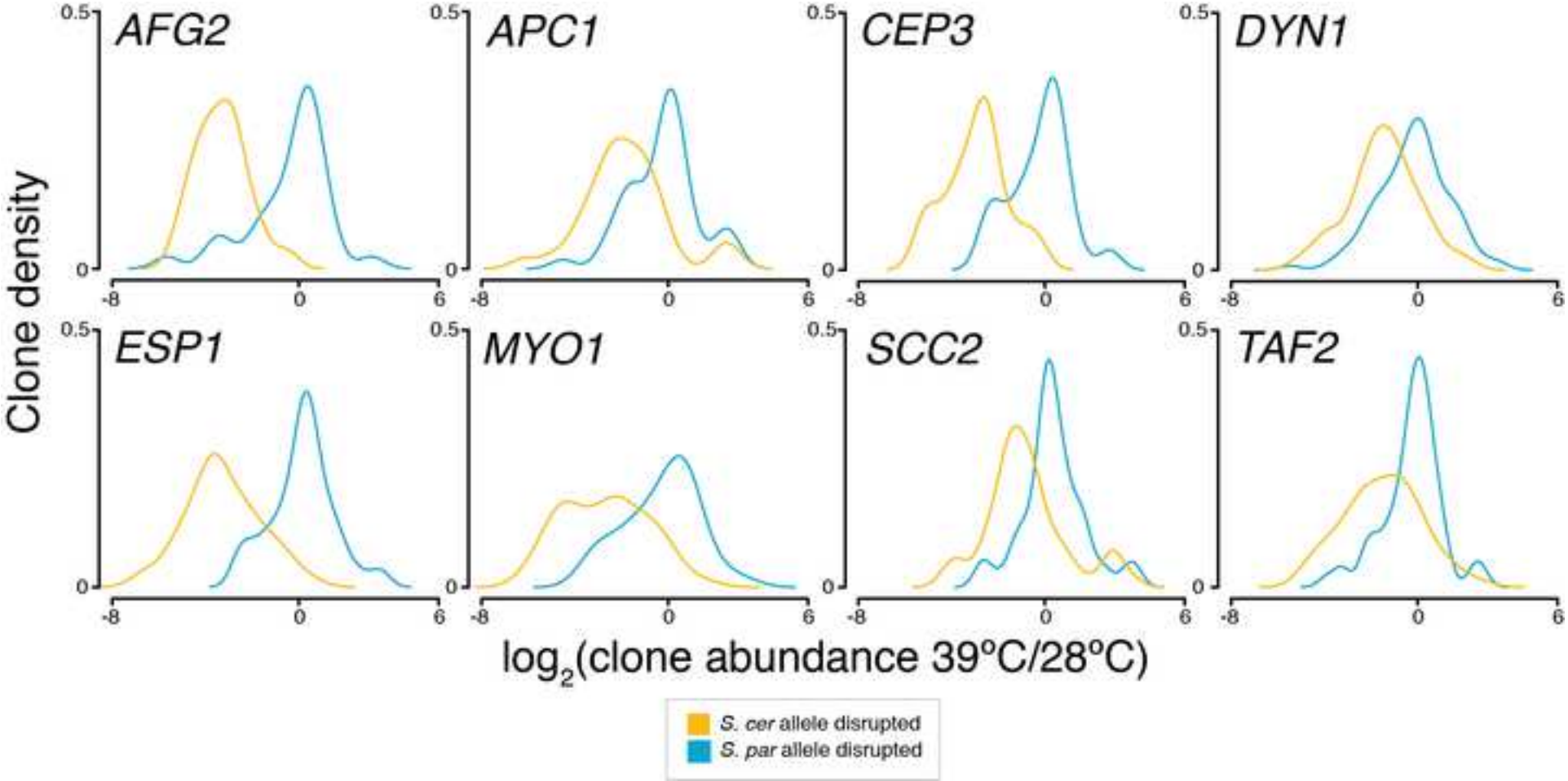


Figure 2

Click here to access/download;Figure;Figure_2.png

Figure 3

Click here to access/download;Figure;Figure_3.png



Click here to access/download;Table of Materials;Weiss_JoVE_table_of_materials_resubmission.xlsx

Name of Material/ Equipment	Company	Catalog Number
1-2 plasmid Gigaprep kits	Zymo Research	D4204
10X Tris-EDTA (TE) buffer (100 mM Tris-HCl and 10 mM EDTA)	Any	N/A
1M LiOAc	Any	N/A
300 mg/mL Geneticin (G418)	Gibco	11811023
52% polyethylene glycol (PEG) 3350	Sigma	1546547
Autoclaved LB liquid broth	BD Difco	244620
Carbenicillin stock in water (100 mg/mL)	Any	N/A
Complete synthetic agar plates (24.1cm x 24.1cm) with 5-fluoroorotic acid (5-FOA) [0.2% drop-out amino acid mix without uracil or yeast nitrogen base (YNB), 0.005% uracil , 2% D-glucose, 0.67% YNB without amino acids, 0.075% 5-FOA]	5-FOA: Zymo Research, Drop-out mix: US Biological, Uracil: Sigma, D-glucose: Sigm), YNB: Difco	5-FOA: F9001-5, Drop-out mix: D9535, Uracil: U0750, D-glucose: G8270, YNB: DF0919
DMSO	Any	N/A
E. coli strain carrying pJR487 (CEN$^-$ / piggyBac-containing plasmid)	N/A	N/A
Hybrid yeast strain JR507 (*S. cerevisiae* DBVPG1373 x *S. paradoxus* Z1, URA$^-$/URA$^-$)	N/A	N/A
Illumina Hiseq 2500		
Large shaking incubators with variable temperature settings	Any	N/A
LB + carbenicillin agar plates (100 µg/mL)	Agar: BD Difco	Agar: 214010

Nanodrop spectrophotometer	Thermo Scientific	ND-2000
Qubit Fluorimeter	Thermo Scientific	Q33240
Salmon sperm DNA	Invitrogen	15632011
Water bath at 39°C	Any	N/A
Yeast fungal gDNA prep kit	Zymo Research	D6005
Yeast peptone dextrose (YPD) liquid media	BD Difco	Peptone: 211677, Yeast Extract: 212750
YPD + G418 agar plates (300 µg/mL)	Agar: BD Difco	Agar: 214010
YPD agar plates	Agar: BD Difco	Agar: 214010

Comments/Description

The number of kits required depends on how efficient your preps are in each kit. This kit comes with 5 individual plasmid prep columns. Run 1 L of saturated E. coli culture through each prep column, as using more than 1 L per column can cause clogging of the prep filter, leading to low yield and poor quality DNA.
Filter sterilize through a 0.22 µm filter before use.
Filter sterilize through a 0.22 µm filter before use.

Dissolve in water and filter sterilize through a 0.22 µm filter before use.
1X trafo mix: 228 uL 52% PEG, 36 uL 1M LiOAc, 36 uL 10X TE buffer
Make LB liquid broth using your powder from any brand, and milliQ water. Autoclave it before use.
Filter sterilize through a 0.22 µm filter before use.

Request from Brem lab.

Request from Brem lab.
used for SE-150 reads

Make LB agar plates as normal and add carbenicillin to 100 µg/mL before drying.

Add filter-sterilized D-glucose to 2% after autoclaving.

Make YPD agar plates as normal and add G418 to 300 µg/mL before drying.

Signed author license agreement

Click here to access/download;Author License Agreement (ALA);Author_License_Agreement_signed.pdf

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article:	Genetic mapping of thermotolerance differences between species of Saccharomyces yeast via genome-wide reciprocal-hemizygosity analysis
Author(s):	Carly V. Weiss and Rachel B. Brem

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☑ Standard Access ☐ Open Access

Item 2: Please select one of the following items:

☑ The Author is **NOT** a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication of the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

CORRESPONDING AUTHOR

Name:	Rachel B. Brem
Department:	
Institution:	Buck Institute for Research on Aging
Title:	Adjunct Professor
Signature:	Rachel Brem
Date:	March 6th, 2019

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Click here to access/download;Rebuttal Letter;Weiss_JoVE_response_to_reviewers.docx

We are delighted that the editor and reviewers found our manuscript suitable in principle for publication in *JoVE*. We have revised the manuscript in line with their suggestions, which we believe has improved it in terms of readability and utility to the community. Below we detail our changes, with the reviewers' original comments in italics and our responses in plain type. We have also reformatted in line with the editor's suggestions, which we do not reproduce here.

Reviewer #1

60-62: Linkage studies have been employed to study species differences in many studies, not just for intraspecific studies. Most studies in fruit flies.

We agree that populations with partial reproductive incompatibility, which have been named as different species, are often used for linkage mapping. The feature of RH-seq that we intended to highlight is that it can be applied to completely incompatible parents. We have revised the introduction to clarify this important point (lines 66-69).

The introduction gives the impression that the RHT has not previously been employed to study species differences. However, the very first study using the RHT (though he didn't call it this) was used to study a difference between Drosophila species (https://www.ncbi.nlm.nih.gov/pubmed/9853753). It may be useful to point readers to a more general review of the RHT, especially for cross-species studies (https://www.ncbi.nlm.nih.gov/pubmed/25278102).
482 - This is incorrect. The test was first applied in 1998 by Stern (https://www.ncbi.nlm.nih.gov/pubmed/9853753) working in Drosophila. The test was "independently rediscovered and named" by Steinmetz in 2002.

We agree and have rectified this important omission. We now cite these practitioners appropriately, including the thorough and compelling Stern review (lines 72-75).

This entire paragraph, or a version of it, should appear in the introduction since most readers will not be familiar with the logic of the RHT.

We agree and now introduce the test in the introduction section of the manuscript (lines 72-88).

532 "have been reproductively isolated for millions of years"

We have made this change.

118 - What is strain JR507? Presumably this is the hybrid strain that you created. You should say this and perhaps say a few words about the methodology and challenges involved in generating hybrid yeast strains.

We now specify the strain origin (lines 135-136).

154 & 157 - Please clarify how to resuspend the yeast. Vortexing or pipetting up and down or gentle mixing?

We now specify vortexing in these steps.

170 - step 2.13 - Shouldn't this solution be prepared a bit earlier?

We now include the making of this solution in an earlier step (lines 177-179).

366 - "for which there are fewer than"

We have made this change.

The mapping protocol appears to require sufficient divergence between DNA sequences to allow unique mapping to one or the other species sequence. This should be noted somewhere.

We agree and now stipulate this in the text (lines 560-564).

374 - "which map to a single location"

We have fixed this typo.

381 - It is not clear why you pool insertions that are within 3bp but not in an identical insertion point. Are these not independent insertions. You should discuss your reasoning more thoroughly.

Given that the PiggyBac transposon has a strong preference for insertion at TTAA sites, we infer that a reads reporting an apparent transposon-genome junction a few base-pairs away from a TTAA is more likely to reflect a sequencing error than an independent insertion at the non-TTAA site. This logic is now spelled out in the manuscript (lines 424-426).

435 - "Each clone resulting from mutagenesis was a hemizygote"

We have now revised this phrase (line 483).

Reviewer #2

1. If possible the authors should include validations by independent allele swaps in purebred parental strains for a few of their top hits.

These data were included in our original research report (Weiss *et al., Nature Genetics* 2018) and are beyond the scope of the current manuscript given that they are not the “representative results” of the RH-seq protocol per se. We now refer a reader to the original paper for these data (lines 502-504).

2. I request authors to expand their Figure 1 to describe all the protocol steps clearly. Especially Figure 1B should be extended to explain the steps involved in the generation of transposon mutants. This edit will significantly enhance the utility of this protocol.

We thank the reviewer for this useful suggestion. We have now added a new Figure 2 with details of the mutagenesis.

Reviewer #3

My main criticism is that while the first part of this protocol is very well documented (generation of transposon mutant library using a hybrid yeast strain and a transposase plasmid), the second part (library generation especially, but also data analysis) is covered in far less detail…The current version of the protocol describes the generation of the transposon mutant library in a lot of detail, but the later steps which are necessary to actually map trait differences are covered in much less detail. For the NGS library generation for Tn-insertion mapping, the protocol just refers to another protocol…As mentioned above I feel it could be advantageous to describe the transposon insertion mapping

(especially NGS library preparation) in more detail.

We thank the reviewer for this in-depth consideration of how to maximize the utility of our manuscript for the community. In our revision we now highlight the published paper whose library generation protocol we used (lines 331-350); this was a key omission in the first submission which we are glad to have rectified. Per the *JoVE* editors, it is in keeping with the journal format guidelines to refer to a published protocol in this scenario rather than reproducing the protocol. As for data analysis, we agree that details of these steps are important. We now have made their description more explicit and added references to the relevant files and Python scripts in our Github repository in the respective steps of the protocol (lines 399, 407, 434, 441, 445, 450, 456, 460, 464, 475).

One could imagine that insertion events in certain non-coding elements could be used for studying their role in trait determination. The possibility of such a 'genome-wide' application could be discussed in more detail…The method described here has some advantages over the method published by Wilkening et al., which could be discussed in more detail.

We agree and have now added mention of this possibility in the discussion (lines 552-553).

The protocol could also be used to characterize haploinsufficiency (in homozygous diploids) or allele-specific effects in heterozygous diploids (analogous to Wilkening et al.).

We agree that the transposon mutagenesis method in principle could be used for these aspects of genomic biology, but we respectfully submit that mention of such an application would be a distraction in the current manuscript, because they are outside the scope of our mapping of species-level trait variation.

The materials listed only cover the steps until generation of the transposon mutant library. As the transposon insertion mapping is an integral part of the protocol in order to dissect trait differences it could make sense to list materials for this part.

We agree that it is critical to make it as easy as possible for readers to implement the Tn-seq library prep. Per the *JoVE* editors as above, the Table of Materials need not list materials for aspects of the protocol that have been published elsewhere.

Recipes for all buffers with amounts in grams or volumes per L medium would be useful to include.

Again, we are grateful to the reviewer for attention to the details of the protocol; in this case we considered that concision of the Table of Materials was a high priority and made the decision to omit these recipes.

For the transformation mix a template table for a 1x mix could be provided.

We have made this addition.

Personally, I would also find it useful to have the steps for generating a hybrid yeast strain (i.e. the input strain for the mutagenesis) briefly explained in the protocol.

We have added these steps to the protocol (lines 135-136).

In line with the comment above, a discussion of how to adapt the protocol for different hybrids with different growth properties (optimal temperature, generation time, optimal transformation parameters)

could be useful, as I imagine different groups would want to perform this protocol for different hybrids. An easy way to achieve this could be to add number of generations for each growth step (in addition to number of hours).

We agree and have added these specifications to the protocol (lines 139, 157, 206, 228, 302).

An example calculation of the expected transformation efficiency (number of Tn events expected, based on the author's experience) would be useful - this would help in adjusting the protocol to a user's specific needs

We refer to this as the transposition efficiency and have added our estimate to the protocol text (lines 268-269).

I am missing a discussion of the problem of cells with >1 insertion. Can the authors provide an estimate of this fraction if one is following their protocol exactly - and/or indicate steps that one can modify to increase/decrease this fraction? Depending on the individual goal a user might want to minimize these events at the cost of getting lower transposition events in the culture, or accept a higher fraction of multiple insertions.

We agree that this is important. In the revised manuscript we now detail what influences the degree to which a given clone could be subject to multiple independent insertions, and how to test it (lines 239-244). We also highlight the robustness of the RH-seq method to noise from secondary mutations in any one clone (lines 549-551).

The authors could add a calculation for sequencing depth - someone without a lot of experience might find this difficult.

In our revised manuscript, we provide an extensive description of the elements of the experiment that govern how much sequencing depth is needed (lines 360-375).

Some advice to improve the fraction of fragments with transposon integration sites over background during library preparation would be useful. The authors mention quite substantial variability of the fraction of Tn events (15% to 90%) among total reads - any indication as to factors that may affect this and how to generally maximize would be good, possibly in a separate troubleshooting section.

This was a typo which we have now rectified (lines 390-391). In fact, our transposon sequencing was very efficient in this study.

Where do the authors get the Tn plasmid from?

As the cloning of the plasmid was described in detail in our earlier research paper (Weiss et al., Nature Genetics 2018) we now give this reference when specifying the plasmid (lines 106-107).

There are only results for 8 top hits in this section. I did not see a supplementary file or link including the count table summarizing the processed data. This is critical for the readers to know how the data should look like and how the calculations were made. And in Figure 2 it is also not clear to me how many inserts and number of reads are taken into account for each of the eight candidate genes.

We agree that the count data are important to give potential users of the protocol an in-depth feel for the outcomes and analysis of RH-seq. The data were published in our original research report (Weiss *et al., Nature Genetics* 2018) which we now refer to explicitly for this purpose (lines 451-454).

Section 3.3. Are 6-7 doublings enough to see differences? Maybe the authors could mention that one could do a culture dilution and grow the pool for more generations (introduces variation but more time for smaller effects to be seen - should be discussed with respect to tradeoff for losing strains with fitness defects).

We thank the reviewer for this important point. In the revised protocol we now discuss the tradeoffs of growing for longer periods, which may help detect smaller effect sizes, but also increases the chance of *de novo* mutations (lines 309-312).

The authors should mention that a high quality reference genome sequence is needed for each of the parental strains - this may not be clear to a less experienced reader.

We agree and have added this point to the revised discussion (lines 562-564).

One step to enrich for 'useful' fragments could be a size selection step after the NGS library preparation to exclude very small fragments (which would not contain enough genomic sequence to unambiguously map the insertion site).

In our revised manuscript we now discuss the size selection inherent in the cleanup steps of the library prep protocol (lines 349-351).

The authors could discuss how hot or cold spots in the genome for Tn mutagenesis could affect the outcome (limitations?). Do they see any evidence of that?

We agree that biased distribution of transposon insertions could limit the utility of the approach. In our original research paper (Weiss *et al., Nature Genetics* 2018) we reported that PiggyBAC achieved a uniform distribution across the genome in our hands, and we have added mention of this in the revised manuscript (lines 410-411).

**5.7 Library size normalization is a common step in analysis of sequencing data. What I suggest here is to use available methods (DESeq2, edgeR) to calculate library sizes. The problem of using the sum /npellet is that it requires the sampling of ninsert to be similar to a normal distribution, which is not always the case for read count data.*

We thank the reviewer for this important potential alternative, which we have added to the protocol in the revised text (lines 433-434).

**5.14 Another advantage of filtering out genes with small insert numbers that could be mentioned is that the phenotypes of these hemizygotes will be less biased by secondary mutations.*

We agree and have stated this explicitly in the revised manuscript (lines 468-470).

**5.15 According to the protocol, replicates are only used to calculate the average abundance, but not used in statistical tests. Additional power might be achieved with a regression model containing all the covariates:*
Fitness (read count) ~ temperature + allele + allele:temperature

The task of the statistical test is then to see if the interaction variable "allele:temperature" explains a good fraction of fitness variation across inserts and replicates of a gene. Existing count data analysis packages like DESeq or EdgeR have implemented such tests in a statistically controlled framework,

with the additional advantage of using the information across replicates and samples to accurately define the variance for each allele.

We agree that this linear modeling framework could well be used as an alternative to our nonparametric Wilcoxon test scheme, and in the revised manuscript we now make this important point (lines 474-475).