

SUPPLEMENTAL INFORMATION: Breaking a species barrier by enabling hybrid recombination

G. Ozan Bozdag, Jasmine Ono, Jai A. Denton, Emre Karakoc, Neil Hunter, Jun-Yi Leu, and Duncan Greig

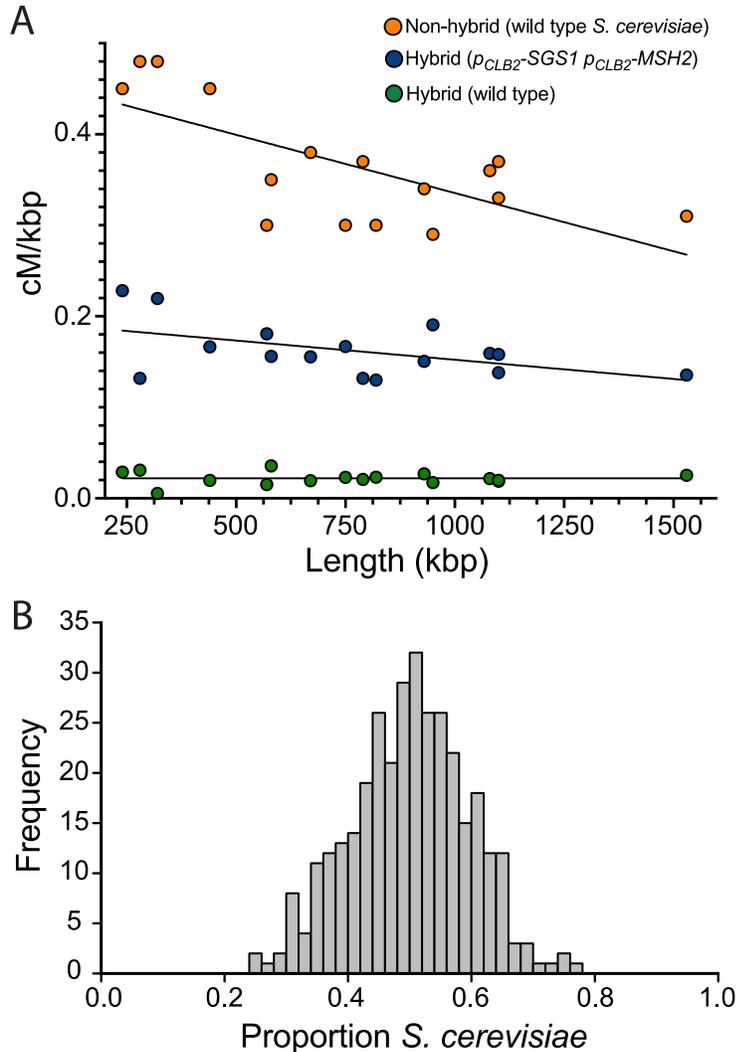


Figure S1: Mutant hybrid yeast show similar patterns of recombination to non-hybrid wild type. **A.** Hybrid double mutants (blue) have overall lower recombination than observed in *S. cerevisiae* (yellow), but much higher recombination than wild type hybrids (green). Similar patterns of more recombination per kbp in shorter chromosomes is observed for both *S. cerevisiae* and the hybrid mutant. Approximate chromosome lengths and genetic map distances were obtained for S288C (SGD)^{S1} and used for all strains even though hybrids are expected to have a mix of chromosome lengths. The hybrid wild type data is from Kao *et al.*^{S2}. **B.** The recombinant spores produced by our manipulated hybrid were normally distributed for inheritance of each species' genome (variance in proportion of the genome inherited from *S. cerevisiae* = 0.0090).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Strains

We used as a template a previously constructed *Saccharomyces cerevisiae* strain NHY 2039, in which the promoter of *SGS1* had been replaced by the *CLB2* promoter^{S3,S4} using the pFA6a-KANMX4pCLB2-3HA construct created by Lee and Amon (2003)^{S5}. We amplified the *S. cerevisiae* *CLB2* promoter and the *KANMX4* drug resistance marker out of NHY2039 (i.e. YDG832) using primer pairs (see below) that allowed us to transform it in place of the natural promoters of *MSH2* and *SGS1* in both *S. cerevisiae* (W303 background) and *S. paradoxus* (N17 background).

The resulting *S. cerevisiae* and *S. paradoxus* haploid strains YDG968 and YDG969 were crossed together producing an F1 hybrid diploid YDG982 in which both homologous copies of both *SGS1* and *MSH2* were under the control of the *CLB2* promoter, repressing the expression of these genes during meiosis (see Data Figure 1 in the Dryad package and strains listed below for details). To obtain a non-hybrid, double-mutant (i.e. *P_{CLB2}-MSH2*, *P_{CLB2}-SGS1*) control strain under the *S. paradoxus* background, we crossed haploid strains YDG967 and YDG968. Next, we crossed YDG969 and YDG970 strains to obtain a similar non-hybrid, double mutant (i.e. *P_{CLB2}-MSH2*, *P_{CLB2}-SGS1*) control strain for the *S. cerevisiae* background. Finally, to obtain a wild-type hybrid control strain (i.e. without *CLB2* promoter replacement), we crossed haploid strains YDG391 (*S. paradoxus*) and YDG542 (*S. cerevisiae*), and selected for diploid clones (to form YDG853).

Fertility

We induced meiosis and sporulation by incubating the hybrid diploid (YDG982) in 3 ml KAc (2% potassium acetate sporulation media) for four days at room temperature with vigorous

shaking. To digest the ascus walls of the hybrid ascospores, we incubated them in 1 unit (per 10 μ l) zymolyase (Zymo Research EU, Freiburg, Germany) for 30 minutes. After enzymatic digestion of the ascus walls, we placed the four spores of each tetrad onto YEPD (2% glucose, 1% yeast extract, 2% peptone, 2% agar) plates using an MSM400 tetrad dissection microscope (Singer Instruments, Watchet, UK). Plates containing dissected tetrads were incubated at 30 °C before examining them for visible colonies founded by germinating spores.

We defined fertility as the proportion of viable gametes, that is, the number of spores that germinated and formed colonies visible to the naked eye after two days, divided by the total number of spores that were dissected. For the hybrid crosses, we dissected a large number of spores (\geq 880, see Data Tables 1 and 2 in the Dryad package). This was necessary for the hybrid crosses because they were known to have low gamete viability. For the non-hybrid crosses, we only dissected 360–400 spores. Because the non-hybrid crosses had much higher rates of gamete viability than the hybrid crosses, dissecting a lower number of spores was sufficient to obtain a good estimate of their true fertility (all standard error < 3%). Only technical replicates (repeated meioses of the same original diploid strain) were performed and they were all considered to be part of a single sample.

Sequencing and genotyping

To ensure that the hybrid gametes we sequenced were euploid, we only genotyped gametes from tetrads that contained four viable spores. In order to maximize useable data from a single lane of sequencing, we limited our sample size to 94 tetrads. Again, repeated meioses of a single diploid strain were performed and were all considered to be part of a single sample. We extracted DNA from all 376 colonies from 94 tetrads (in addition to two non-hybrid control tetrads) using MasterPure™ Yeast DNA Purification Kit (Epicentre, Biozyme Biotech, Oldendorf, Germany).

To prepare the samples for sequencing, we used double digestion based RAD-tag library preparation method^{S6,S7,S8}. We digested 50 ng of DNA from each colony using restriction enzymes *Csp6I* and *PstI* and ligated adapters (adapterX_TagY_fq and adapterX_TagY_rv) in the same reaction at 37 °C for two hours. We cleaned up the excess adapters, enzymes, and fragments smaller than 300bp by using Ampure beads at a 1:1 ratio. Next we mixed Phusion Hot Start II High-Fidelity DNA Polymerase (2U/μl), adding P5 and P7 primers at 10 mM concentration, dNTPs (2mM per dNTP), and 5X Phusion HF Buffer to amplify the target regions^{S6,S9}. 30 μl PCR mixtures were amplified using an initial 98°C incubation for 30s, followed by 25 cycles of 98°C for 10s, 68°C for 15s, 72°C for 30s, and then a final extension at 72°C for 5 mins. To sequence the tagged samples, we mixed all tagged samples in one pool. All samples were multiplexed using combinations of 24 unique barcodes, therefore reads from a single sequencing reaction had unique reverse and forward tags, which allowed us to distinguish the samples. We used the MiSeq platform to obtain 300 bp paired-end reads.

To map the reads, we assembled two simplified collinear reference genomes consisting of the coding DNA only from the set of open reading frames (ORFs) shared between *S. cerevisiae* and *S. paradoxus*, removing ORFs that were present in one species but not in another or which were not collinear (based on SGRP sequencing added 10/10/08^{S10,S11}). We mapped reads to these reference genomes using bwa-mem^{S12}. At this point, we excluded 10 tetrads due to poor sequencing coverage and quality, leaving us with 336 samples from 84 tetrads. We assigned ORFs to one species or the other using two simplifying assumptions: that no non-Mendelian segregation occurred and that recombination occurred only in intergenic regions. Thus, if all four spores of a tetrad contained reads mapping to a given ORF of one or both species, the two spores with the highest proportion of reads mapping to one species' ORF would have it assigned to that species and the other two would have the ORF assigned to the other species. If the four spores

within a tetrad did not all contain reads for an ORF, it would be assigned to the same species as the neighboring ORF. These genotyping rules produced a recombination map (see Figure 1C for an example, full data available from Dryad) of the four spores within each tetrad at ORF-level resolution, with no gain or loss of genetic material (that is, no gene conversion). To visualize the genotype of these hybrid spores, we used PlotTetradSeq in the Recombine package in R^{S13}.

Statistics

χ^2 contingency tests for viability statistics and comparing the production of four-spore tetrads between strains were performed using Pearson's χ^2 test with Yates' continuity correction using the function 'chisq.test' from the package stats in R^{S14}. The standard error for the viability data was calculated manually in R using the formula $SE = \sqrt{p(1-p)/n}$. The 95% confidence interval was calculated using the Agresti-Coull method in the function 'binom.confint' from the package binom in R^{S14}. Linear regressions looking at each strain's cM/kbp were performed using the 'linregress' function from the scipy.stats package in Python^{S15}, which fits a linear regression line that minimizes the sum-of-squares for two sets of measurements.

Primers used in this study

To replace the native promoter of *SGS1* with *KANMX-pCLB2SGS1* cassette (can be amplified from YDG832 or YDG833):

S. cerevisiae

FkMXpClb2Sgs1Sc

ATTATTGTTGTATATATTTAAAAAATCATAACACGTACACACAAGGCGGTAGAATTCG
AGCTCGTTTAAAC

RkMXpClb2Sgs1Sc

TTTAACCATTTGTGCTCCCTTCTTAAGTTATGTGACGGCTTCGTCACCATGCACTGAG
CAGCGTAATCT

S. paradoxus

FkMXpCib2Sgs1Sp

AGTTCAGTGTATATATTTAAAGTCACACGCATACACGCGAATTCGAGCTCGTTTAA
C

or

SpKMXpCLB2sgs1F

CAGCCAGTTTGGAGTCATCAGCTACAGGAAGGGAAAACGAATTCGAGCTCGTTTAA
AC

RkMXpCib2Sgs1Sp

TTCAACCATTTGTGCTCCCTTCTTAAGTTATGTGAGGGCTTCGTCACCATGCACTGAG
CAGCGTAATCT

To diagnose replacement of native *SGS1* promoter with *KANMX-pCLB2* cassette:

RdKanMX

GTTCGGATGTGATGTGAGAACTG

RdSgs1orfSc

TGTGCTTTTGGATAGCCTGG

To replace the native promoter of *MSH2* with *KANMX-pCLB2* cassette:

S. cerevisiae

FpCLBmsh2c

TTATCTGCTGACCTAACATCAAAATCCTCAGATTAAGTGAATTCGAGCTCGTTTAA
AAC

RpCLBmsh2c

ATACATCAGAGAATTTTAGCTCTGGCCTAGTGGAGGACATGCACTGAGCAGCGTAAT
CT

S. paradoxus

FpCLBmsh2p

TTATCTGTTGATCTAACATCAAAATCTTCGGATAACAAGTGAATTCGAGCTCGTTTAA
AAC

RpCLBmsh2p

ATACATCAGAGAATTTTCAGCTCTGGTCTAGTAGAGGACATGCACTGAGCAGCGTAAT
CT

To diagnose replacement of native *MSH2* promoter with *KANMX-pCLB2* cassette:

S. cerevisiae (used with any forward primer that anneals to *KANMX* cassette).

Rd-pCLBmsh2c

CATAACTTCAGCAGAGTGGC

S. paradoxus

Rd-pCLBmsh2p

ATAATTCAGCAGAGTGG

Strains used in this study

YDG227: *ho MATa lys2 cyh2r* (N17)

YDG244: *ho MATalpha ura3 cyh2r* (N17)

YDG253: *ho MATa ura lys his3 leu2 trp ade-2can1r* (W303)

YDG350: *ho MATalpha ura3-52 his3-11 leu2-3,112 trp1Δ2 ade2-1 can1-100* (W303, YSC1059)

YDG391: *ho::HYGMX MATalpha ura3::KanMX* (N17, NCYC 3708)

YDG542: *ho::HYGMX MATa ura3::KanMX ade2-1* (W303, NCYC 3583)

YDG832: *ho::hisG MATa ura3(ΔSma-Pst) HIS4::LEU2-(BamHI; +ori) leu2::hisG pCLB2-3HA-SGS1::kanMX4* (NHY 2039)

YDG833: *ho::hisG MATalpha ura3(ΔSma-Pst) his4-X::LEU2-(NgoMIV; +ori)—URA3 leu2::hisG pCLB2-3HA-SGS1::kanMX4* (NHY 2040)

YDG853: *ho::HYGMX/ho::HYGMX MATa/MATalpha ura3::KanMX/ura3::KanMX ade2-1/ADE2* (N17 x W303, YDG391 x YDG542)

YDG863: *ho MATa ura lys his3 leu2 trp ade-2 can1r pCLB2-3HA-SGS1::kanMX4* (W303)

YDG866: *ho MATa ura his3 leu2::NAT trp ade-2 can1r pCLB2-3HA-SGS1::kanMX4* (W303)

YDG905: *ho MATalpha ura3 cyh2r pCLB2-3HA-SGS1::kanMX4* (N17)

YDG912: *ho/ho MATa/MATalpha ura/ura3 his3/HIS3 leu2::NAT/LEU2 trp/TRP ade-2/ADE can1r/CAN1 CYH2/cyh2r pCLB2-3HA-SGS1::kanMX4/pCLB2-3HA-SGS1::kanMX4* (W303 x N17, YDG866 x YDG905)

YDG959: *ho MATa lys2 cyh2r pCLB2-3HA-MSH2::kanMX4* (N17)

YDG960: *ho MATalpha ura3-52 his3-11 leu2-3,112 trp1Δ2 ade2-1 can1-100 pCLB2-3HA-MSH2::kanMX4* (W303, YSC1059)

YDG961: *ho MATa lys2 cyh2r pCLB2-3HA-SGS1::kanMX4* (N17)

YDG962: *ho MATalpha ura3-52 his3-11 leu2-3,112 trp1Δ2 ade2-1 can1-100 pCLB2-3HA-SGS1::kanMX4* (W303, YSC1059)

YDG964: *ho/ho MATa/MATalpha URA3/ura3-52 lys2/LYS2 HIS3/his3-11 LEU2/leu2-3,112 TRP1/trp1Δ2 ADE2/ade2-1 CAN1/can1-100 cyh2r/CYH2 pCLB2-3HA-MSH2::kanMX4/pCLB2-3HA-MSH2::kanMX4* (N17 x W303, YDG959 x YDG960)

YDG965: *ho/ho MATalpha/MATa ura3/URA3 LYS2/lys2 cyh2r/cyh2r pCLB2-3HA-SGS1::kanMX4/pCLB2-3HA-SGS1::kanMX4* (N17, YDG905 x YDG961)

YDG967: *ho MATa cyh2r pCLB2-3HA-SGS1::kanMX4 pCLB2-3HA-MSH2::kanMX4* (N17)

YDG968: *ho MATalpha cyh2r pCLB2-3HA-SGS1::kanMX4 pCLB2-3HA-MSH2::kanMX4* (N17)

YDG969: *ho MATa ura3 his3 leu2::NAT trp1 ade-2 can1r pCLB2-3HA-SGS1::kanMX4 pCLB2-3HA-MSH2::kanMX4* (W303)

YDG970: *ho MATalpha ura3 his3 leu2::NAT trp1 ade-2 can1r pCLB2-3HA-SGS1::kanMX4 pCLB2-3HA-MSH2::kanMX4* (W303)

YDG979: *ho/ho MATa/MATalpha ura/ura3-52 his3/his3-11 leu2::NAT/leu2-3,112 trp/trp1Δ2 ade-2/ade-2-1 can1r/can1-100 pCLB2-3HA-SGS1::kanMX4/pCLB2-3HA-SGS1::kanMX4* (W303, YDG866 x YDG962)

YDG980: *ho/ho MATa/MATalpha cyh2r/cyh2r pCLB2-3HA-SGS1::kanMX4/pCLB2-3HA-SGS1::kanMX4 pCLB2-3HA-MSH2::kanMX4/pCLB2-3HA-MSH2::kanMX4* (N17, YDG967 x YDG968)

YDG981: *ho/ho MATa/MATalpha ura3/ura3 his3/his3 leu2::NAT/leu2::NAT trp1/trp1 ade-2/ade-2 can1r/can1r pCLB2-3HA-SGS1::kanMX4/pCLB2-3HA-SGS1::kanMX4 pCLB2-3HA-MSH2::kanMX4/pCLB2-3HA-MSH2::kanMX4 (W303, YDG969 x YDG970)*

YDG982: *ho/ho MATalpha/MATa URA3/ura3 HIS3/his3 LEU2/leu2::NAT TRP1/trp1 ADE/ade-2 CAN1/can1r cyh2r/CYH2 pCLB2-3HA-SGS1::kanMX4/pCLB2-3HA-SGS1::kanMX4 pCLB2-3HA-MSH2::kanMX4/pCLB2-3HA-MSH2::kanMX4 (N17 x W303, YDG968 x YDG969)*

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jasmine Ono (j.ono@ucl.ac.uk).

Materials Availability

Strains generated in this study are available on request.

Data and Code Availability

The datasets generated during the current study have been deposited in the Dryad repository, doi:10.5061/dryad.v41ns1rtn. Source data for Figure 1A and B is provided in the Dryad package as Data Tables 1 and 2. Source data for recombination rates and Figure S1A is provided in Data Table 3 of the Dryad package. Figure 1C and Figure S1B can be generated from data in Data Files 1 and 2, located in the Dryad package. The raw sequence data that support the findings of this study have been deposited in NCBI (SRA accession number: PRJNA672474).

AUTHOR CONTRIBUTIONS

G.O.B., D.G., J.Y.-L., and J.O. conceived and directed the project. D.G. and J.Y.-L. acquired the funding. D.G. provided the laboratory resources and oversaw the project. G.O.B., J.O., J.A.D., N.H., and D.G. designed the methods employed. G.O.B., J.A.D., and D.G. coordinated the research. G.O.B. performed the data collection. G.O.B., J.O., E.K., and D.G. contributed to the statistical analyses. G.O.B., J.O., and E.K. wrote the code to perform the analyses and verified the results. G.O.B. and J.O. prepared the visualisations and performed the data curation. J.O. and D.G. wrote the original draft, G.O.B. edited the manuscript, and N.H. reviewed the manuscript.

SUPPLEMENTAL REFERENCES

- S1. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., *et al.* (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* *40*, D700–D705.
- S2. Kao, K.C., Schwartz, K., and Sherlock, G. (2010). A genome-wide analysis reveals no nuclear Dobzhansky-Muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. *PLoS Genet.* *6*, e1001038.
- S3. Grandin, N., and Reed, S.I. (1993). Differential function and expression of *Saccharomyces cerevisiae* B-type cyclins in mitosis and meiosis. *Mol. Cell. Biol.* *13*, 2113–2125.
- S4. Oh, S.D., Lao, J.P., Taylor, A.F., Smith, G.R., and Hunter, N. (2008). RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. *Mol. Cell* *31*, 324–336.
- S5. Lee, B.H., and Amon, A. (2003). Role of Polo-like kinase CDC5 in programming meiosis I chromosome segregation. *Science* *300*, 482–486.
- S6. Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., and Cresko, W.A. (2012). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In *Molecular*

Methods for Evolutionary Genetics, V. Orgogozo and M.V. Rockman, ed. (Springer), pp. 157–178.

- S7. Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., and Cresko, W.A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6, e1000862.
- S8. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7, e37135.
- S9. Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* 71, 8966–8969.
- S10. Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., and Koufopanou, V. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341.
- S11. Bergström, A., Simpson, J.T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A.N., Moses, A.M., Louis, E.J., and Mustonen, V. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* 31, 872–888.
- S12. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, arXiv:1303.3997.
- S13. Anderson, C.M., Chen, S.Y., Dimon, M.T., Oke, A., DeRisi, J.L., and Fung, J.C. (2011). ReCombine: a suite of programs for detection and analysis of meiotic recombination in whole-genome datasets. *PLoS One* 6, e25509.
- S14. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- S15. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., *et al.* (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272.