

Supplementary Materials for
**Non-paradoxical evolutionary stability of the recombination initiation
landscape in yeast**

Isabel Lam^{1,2} and Scott Keeney^{1,2,3,*}

¹ Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA.

² Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA.

³ Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA.

*Correspondence to: s-keeney@ski.mskcc.org.

This PDF file includes:

Materials and Methods
Figures S1-S6
Tables S1-S2

Other Supplementary Materials for this manuscript includes the following:

Tables S3-S5 (provided as separate Excel files):

Table S3. Hotspot lists in *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*.

Table S4. List of 3426 matched promoter intergenic regions.

Table S5. List of 20-kb bins for large-scale interstitial analysis.

Materials and Methods:

Yeast strains and culture methods

Studies were done in the following *Saccharomyces* species: *S. cerevisiae* SK1, YPS128 and UWOPS03-461.4, *S. paradoxus* YPS138, *S. mikatae* IFO1815, *S. kudriavzevii* ZP591 (**Table S1**). *SPO11* in each species was C-terminally tagged with three Flag epitope repeats by targeted integration of a *6His-3FLAG-loxP-kanMX-loxP* construct amplified from an *S. cerevisiae* SK1 *SPO11-Flag* strain provided by Kunihiro Ohta, Univ. Tokyo (43)). Yeasts were transformed using standard lithium acetate methods, with modifications to heat shock incubation times and temperatures as follows. Heat shocks were done for 5 min at 37°C for *S. mikatae* and 30 min at 34°C for *S. kudriavzevii*, as described (44). For *S. paradoxus*, heat shock was for 8 min at 37°C, on the basis of advice generously provided by Jeremy Roop (Rachel Brem laboratory, UC Berkeley). Since all strains were homothallic diploids, homozygous tagged strains were generated by transformation, followed by sporulation, tetrad dissection, and screening self-diploidized spore clones for resistance to G418. Correct tagging was verified by PCR and Southern blot.

Synchronous meiotic cultures of *S. cerevisiae* SK1 were prepared as described in (45), with 14 h pre-sporulation in YPA (1% yeast extract, 2% peptone, 1% potassium acetate) and sporulation in 2% potassium acetate, 0.2 × supplements. For wild-derived *S. cerevisiae*, *S. paradoxus* and *S. kudriavzevii*, 14 h pre-sporulation in YPA was followed by sporulation in 1% potassium acetate, 0.2 × supplements. For *S. mikatae*, the above conditions result in premature meiotic entry, so sporulation conditions were based instead on methods of (46). Briefly, a 7 h starter culture in SPS (0.5% yeast extract, 1% peptone, 0.67% yeast nitrogen base without amino acids, 1% potassium acetate, 0.05 M potassium biphthalate, pH 5.5) was used to inoculate a 16 h pre-sporulation culture in SPS, and sporulation was done in 1% potassium acetate, 0.001% polypropylene glycol supplemented with 0.32% amino acid complementation medium (1.5% lysine, 2% histidine, 2% arginine, 1% leucine, 0.2% uracil, 1% tryptophan). All SK1 culturing steps were performed at 30°C. All other strains and species were cultivated at room temperature (growth on solid media) or 23°C (all growth in liquid media). For all strains and species, cells were transferred to sporulation media to a cell density (OD₆₀₀) of 6.0. For Spo11-oligo mapping, ≥ 600 ml sporulation culture volume was harvested 4 h (*S. cerevisiae* YPS128, *S. paradoxus*, *S. mikatae*), 6 h (*S. cerevisiae* UWOPS03-461.4), or 9 h (*S. kudriavzevii*) after transfer to sporulation media, corresponding with the approximate time of peak levels of Spo11-oligo complexes (**fig. S1B**).

To assess culture synchrony, meiotic division profiles were obtained by collecting aliquots at various times from synchronous meiotic cultures, fixing in 50% (v/v) ethanol, and staining with 0.05 µg/ml 4', 6-diamidino-2-phenylindole (DAPI). Mono-, bi- and tetranucleate cells were scored by fluorescence microscopy.

End-labeling of Spo11-oligo complexes and Spo11-oligo mapping

To detect Spo11-oligo complexes, previously described methods were used for denaturing cell extract preparation, anti-Flag immunoprecipitation (IP), end-labeling with terminal deoxynucleotidyl transferase, and SDS-PAGE (32, 45). Radiolabeled complexes were detected by phosphorimager.

For SK1, we used a previously published Spo11-oligo map (two biological replicates) prepared from a strain in which Spo11 was tagged with five copies of a fragment of protein A

(32). New maps were generated for this study in the other strains/species carrying *SPO11-Flag*. Processing of cell lysates for Spo11-oligo purification and sequencing library preparation were essentially as described (47), except that Spo11-Flag IP was carried out with protein G Dynabeads (Life Technologies) instead of protein G agarose beads. In the first round of IP, 400 μ l protein G Dynabeads pre-bound with 80 μ l of 1 mg/ml mouse anti-Flag M2 antibody (Sigma) was used per 25 ml whole-cell extract in 50 ml IP volume. Pre-binding of antibody to beads was carried out with pre-washed protein G Dynabeads (washed twice with 1 \times IP buffer (1% Triton X-100, 15 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA)) and anti-Flag antibody in 1 ml 1 \times IP buffer for 10 min at 37°C, with end-over-end rotation. Beads were subsequently washed twice with 1 \times IP buffer prior to use in the IP. For the second round of IP, 125 μ l protein G Dynabeads pre-bound with 25 μ l of 1 mg/ml anti-Flag antibody was used in an 800 μ l IP volume. Sequencing (Illumina HiSeq 2500, 2 x 75 bp paired-end reads) was performed in the Integrated Genomics Operation at MSKCC.

Clipping of library adapters and mapping of reads was performed by the Bioinformatics Core Facility (MSKCC) using a custom pipeline as described (15, 32, 47). *S. cerevisiae* SK1, YPS128, and UWOPS03-461.4 Spo11-oligo reads were mapped to the sacCer2 genome assembly of type strain S288C from SGD (*Saccharomyces* Genome Database), *S. paradoxus* reads were mapped to *S. paradoxus* YPS138 genome assembly from SGRP (*Saccharomyces* Genome Resequencing Project) or CBS432 type strain genome assembly from (44), and *S. mikatae* and *S. kudriavzevii* were mapped to their strain genome assemblies from (44). The genome assemblies of *S. mikatae* and *S. kudriavzevii* contain unplaced contigs; mapping included these contigs, but downstream analyses were done on chromosomes 1–16 only. For all analyses of *S. paradoxus*, we used maps based on the genome assembly for the same strain background from which Spo11 oligos were purified (YPS138), except for the analysis of large-scale correlations with GC content in **Fig. 4E**, in which Spo11-oligo reads mapped to the type strain CBS432 were compared with GC content in the CBS432 genome sequence due to the latter being a more complete sequence assembly. Sequence read totals and mapping statistics are described in **Table S2**. Statistical analyses were performed using the R program package (RStudio version 0.98.1091, R version 3.0.1).

Raw and processed sequence reads have been deposited in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) as accession number GSE71887. This accession also contains the curated maps (unique mapping reads only, normalized to reads per million mapped) in wiggle format to allow direct visualization in appropriate genome browsers, e.g., the UCSC browser (<https://genome.ucsc.edu>) using genome version sacCer2 for *S. cerevisiae* maps, or the Integrative Genomics Viewer (IGV) genome browser (Broad Institute) and loading the appropriate genome assembly files.

DSB hotspots were identified using the method from (15), with minor modifications. Hotspots were defined as clusters of Spo11 oligos meeting cutoffs for size and Spo11-oligo density similar to the previous definitions. Briefly, candidate hotspots were first identified as regions where the Spo11-oligo map smoothed with a 201-bp Hann window was >0.193 RPM per bp (which is ≥ 2.2 -fold over the genome average Spo11-oligo density in the four species; more specifically: 2.3-fold in *S. cerevisiae* and *S. paradoxus*, 2.2-fold in *S. mikatae* and *S. kudriavzevii*). Within these candidate regions, hotspot boundaries were defined as the leftmost and rightmost coordinates to which reads were mapped. Hotspots separated by less than 200 bp were merged, then hotspot lists were filtered to remove calls less than 25 bp wide and/or containing fewer than 10 RPM total.

Defining matched promoter-containing IGRs

To evaluate evolutionary conservation of hotspot heats, we used an IGR-centric approach rather than one based on the more arbitrarily defined hotspots for several reasons. First, the substantial sequence divergence within IGRs (where most hotspots lie) makes it difficult or impossible to define precisely where a hotspot boundary identified in one species lies in the genomic space of another species. In contrast, IGRs are generally unambiguous because they are defined by the much better conserved coding sequence of the flanking genes. Second, hotspot definitions are highly arbitrary: the smooth continuum of hotspot heats (**Fig. 2E**, **fig. S3E**) means that there is no biologically defined cutoff between what is and what is not a hotspot (15). Third, by using IGRs rather than called hotspots, we can include matched DSB-cold promoter regions on the same footing as DSB-hot ones.

Lists of IGRs were compiled for *S. cerevisiae* S288C (sacCer2 assembly), *S. paradoxus* YPS138 (SGRP), *S. mikatae* IFO1815 (44), and *S. kudriavzevii* ZP591 (44) based on the list of genes in their respective genome annotations. Genes annotated as “dubious ORFs” in *S. cerevisiae* were excluded in all species prior to compiling IGRs. After further excluding overlapping genes/transcription units, total numbers of IGRs and the breakdown into divergent-, tandem-, and convergent-oriented flanking transcription units compiled were: *S. cerevisiae*, 5715 (1493 divergent, 2746 tandem, 1476 convergent); *S. paradoxus*, 5337 (1404 divergent, 2539 tandem, 1394 convergent); *S. mikatae*, 5637 (1458 divergent, 2726 tandem, 1453 convergent); *S. kudriavzevii*, 5549 (1436 divergent, 2678 tandem, 1435 convergent).

IGRs were assigned names by concatenating the annotated names of the flanking genes. A list (n=7138) of the names of all IGRs present in the four species was then compiled with unique IGR names as separate entries (that is, IGR names found in multiple species were entered only once). If the assigned IGR name was the same in all four species, the IGR was provisionally designated as matched. IGRs not matched this way were then manually curated. Oftentimes, apparently species-specific IGRs simply reflected mis-annotation of one flanking gene with the name of a paralog or of another member of a multi-gene family, so these were changed to match the gene name order in *S. cerevisiae* and scored as matched. Annotations of IGRs close to the chromosome ends that were different from *S. cerevisiae* were left as is, since subtelomeric region sequences pose a challenge for whole-genome assembly due to duplication blocks and repeats (29), and tend to be less conserved between yeast species (48). IGR names missing in a species due to indels or gaps in the genome sequence were also left as is.

After this curation, the list contained 6656 different IGR names among the four species. The following criteria were then used to define matched promoter-containing IGRs: the IGR name is present in all four species, is on the same chromosome (except if the IGR is on a segment that had undergone reciprocal translocation in *S. mikatae*), has the same flanking transcript orientation, and contains a promoter(s) (i.e., divergent or tandem orientation) in all four species. After applying the above criteria and excluding IGRs with no sequence coverage, the list was narrowed down to 3426 matched promoter-containing IGRs (**Table S4**).

To calculate the fractional overlap of hotspots between species within the defined regions of synteny encompassing the 3426 promoter IGRs, the ~4000 hotspots from each species were queried for overlap with the coordinates of the IGRs. Approximately 70% of total hotspots in each species overlapped the 3426 promoter IGRs based on this criterion (71.7% in *S. cerevisiae* SK1, 72.0% YPS128, 71.4% UWOPS03-461.4, 74.2% *S. paradoxus*, 73.2% *S. mikatae*, 70.2% *S. kudriavzevii*). The remaining ~30% of hotspots in each species that are excluded from

subsequent analyses either overlap promoter IGRs that are not in common amongst the four species based on our stringent criteria for inter-species IGR matching (e.g., 19% of *S. cerevisiae* promoter IGRs), or do not overlap IGRs.

Percent sequence divergence was calculated for the 3426 matched promoter IGRs by performing global (Needleman-Wunsch) pairwise sequence alignments (`pairwiseAlignment` function from Biostrings package) for each IGR sequence in all six pairwise inter-species comparisons, taking the median percent sequence identity across the 3426 IGR sequence alignments, and subtracting from 100%. Genome assembly and annotations for *S. cerevisiae* in the inter-species comparisons were from strain S288C (SGD sacCer2).

For intra-species (SK1, YPS128, UWOPS03-461.4) sequence divergence calculations, genome assemblies and annotations for each *S. cerevisiae* strain were from SGRP. Because of incomplete genome assemblies and annotations, fewer matched promoter IGRs could be aligned in the pairwise intra-species comparisons. For example, IGR sequences with stretches of “N”s (representing unknown sequence) longer than 25% of the IGR length, instances where the IGR coordinates extend beyond the length of the chromosome-length scaffold, or instances where the IGR name did not match any of the IGRs in the strain’s genome annotation were excluded from the pairwise alignment analysis. Total numbers of IGRs aligned are 3368 in SK1 vs. UWOPS03-461.4 and SK1 vs. YPS128 (out of 3426, or 98.3% of matched promoter IGRs), and 3385 in UWOPS03-461.4 vs. YPS128 (98.8%).

Analysis of large-scale interstitial regions

To compare syntenic interstitial genomic regions between species, a common coordinate system was generated for pairwise comparisons with *S. cerevisiae* by using syntenic genes as points of reference when dividing the genome into bins. First, stretches of synteny within interstitial regions (excluding 20 kb from ends of chromosomes, and 10 kb from each side of the centromere) were defined by going through the list of genes in the *S. cerevisiae* S288C genome annotation, and querying whether there was a match with the same gene order in the species being compared (i.e., the gene was consecutive to the previous gene that had a match). If there was no match for a gene in the appropriate position in the second species, the stretch of synteny ended, and a new one was begun. Although the gene lists were curated for mis-annotations with paralog/multi-gene family names to maximize synteny with *S. cerevisiae* (see above), many of the remaining breaks in synteny probably reflect unresolved annotation discrepancies. Other synteny breaks likely reflect species or strain-specific presence of transposable elements or copy number variants for multicopy gene families.

Bins of 20 kb were subsequently defined in *S. cerevisiae* within each stretch of synteny, starting with the left-most base pair coordinate in the stretch of synteny, and calculating the midpoint (start + 10 kb) and end coordinates of the bin. To determine the equivalent bin coordinates in the comparand species, we measured the distance from the midpoint to the start of the next gene in *S. cerevisiae*. The start coordinate of the orthologous gene in the comparand species was then used as a fiduciary mark, and the bin midpoint was defined as lying the same distance from this mark as was measured in *S. cerevisiae*. Bin boundaries were then set 10 kb to the left and right of this midpoint, and Spo11 oligos were summed within the bins. Total numbers of 20-kb bins defined genome-wide in each species’ comparison with *S. cerevisiae* are: *S. paradoxus*, 491; *S. mikatae*, 448; *S. kudriavzevii*, 459. Spo11-oligo counts were then summed up within the start and end coordinates of each 20-kb bin in each species. Bin coordinates and Spo11-oligo counts are provided in **Table S5**.

Coordinates of *S. mikatae* translocation

The two reciprocal translocations in *S. mikatae* IFO1815 were originally mapped by pulsed-field gel electrophoresis (49), and later confirmed with whole-genome sequencing (22, 44). Based on the gene order in the *Saccharomyces sensu stricto* Resources genome browser (www.saccharomycessensustricto.org) (44), the translocation breakpoints can be further narrowed down to chr VI *YFR006W*-*YFR009W*, chr VII *YGR023W*-*YGR027C* and *YGR188C*-*YGR189W*, and chr XVI *YPL116W*-*YPL103C*. Thus, the left-most ~165 kb region of *S. mikatae* chr VI, which encompasses the left arm and centromere, is syntenic with the left arm and centromere of *S. cerevisiae* chr VI up to *YFR006W*, then transitions to a ~303 kb segment syntenic with *S. cerevisiae* chr VII from *YGR027C* to *YGR187C*, and switches to a ~308 kb segment syntenic with *S. cerevisiae* chr XVI from *YPL116W* to *YPL273W*. The left-most ~531 kb of *S. mikatae* chr VII is syntenic with *S. cerevisiae*, up to *YGR023W*, then transitions to a ~110 kb segment syntenic with the right arm of *S. cerevisiae* chr VI (starting with *YFR009W*). The left arm of *S. mikatae* chr XVI is syntenic with a ~193 kb segment in *S. cerevisiae* chr VII, from *YGR288W* to *YGR189W*, but the remainder of the left arm (starting from *YPL103C*), centromere, and right arm are syntenic with the *S. cerevisiae* chr XVI.

To calculate Spo11-oligo density within syntenic segments on different chromosome-length contexts in *S. cerevisiae* and *S. mikatae*, the following coordinates were used: *S. cerevisiae* 6L, chr VI 1–160,000 bp; *S. mikatae* 6L, chr VI 1–165,160 bp (sequence to the left of *YFR006W*/Smik_6.88); *S. cerevisiae* 6R, chr VI 162,000–270,161 bp; *S. mikatae* 7R, chr VII 524,000–628,517 bp (sequence to the right of *YFR009W*/Smik_7.313). Coordinates for control segments are as follow: *S. cerevisiae* 4L, chr IV 22,823–190,586 bp; *S. mikatae* 4L, chr IV 887–167,444 bp (*YKL124W*/Smik_11.115 to *YDL148C*/Smik_4.87); *S. cerevisiae* 11L, chr XI 210,093–371,390 bp; *S. mikatae* 11L, chr XI 196,553–357,829 bp (*YKL124W*/Smik_11.115 to *YKL035W*/Smik_11.212).

Nucleosome occupancy (MNase-seq)

Nucleosome mapping in the *Saccharomyces* species was performed as described for SK1 meiotic nucleosome maps in (15). Synchronous meiotic cultures were set up for each species and/or strain background as described above, except volumes were scaled-down to 200 ml for pre-sporulation, and approximately 110 ml for sporulation, at $OD_{600} = 1.9$. 100 ml samples were harvested at 4 h (*S. paradoxus*, *S. mikatae*, and *S. cerevisiae* YPS128), 6 h (*S. cerevisiae* UWOPS03-461.4), or 9 h (*S. kudriavzevii*) after transfer to sporulation media, and cross-linked with 1% formaldehyde for 15 min with rocking. Chromatin extraction, MNase digestion, DNA purification, and sequencing library construction were performed as described (15). Amounts of MNase used to generate mononucleosome-sized DNA were 5 U and 10 U (*S. mikatae*); 10 U (*S. cerevisiae* UWOPS03-461.4); 20 U (*S. kudriavzevii*); 20 U and 40 U (*S. cerevisiae* YPS128, *S. paradoxus*). Library preparation and Illumina sequencing (2 x 50 bp paired-end reads) were performed by the MSKCC Integrated Genomics Operation.

S. cerevisiae YPS128 and UWOPS03-461.4. Illumina reads were mapped to the sacCer2 genome assembly, *S. paradoxus* reads were mapped to *S. paradoxus* YPS138 genome assembly from SGRP or CBS432 type strain genome assembly from (44), and *S. mikatae* and *S. kudriavzevii* were mapped to their strain genome assemblies from (44). The reads were first clipped to remove any Illumina adapter sequences present and reads shorter than 35 bp were discarded (both ends of paired end reads were discarded). The clipped reads were then mapped to

the proper target genome using BWA MEM (default options). The output SAM files were coordinate sorted and read groups added and merged into single sample BAM files using the PICARD toolkit. The output BAM files were then post-processed with a series of custom scripts to filter for uniquely mapping and properly paired reads. Proper pairing was defined as reads where the chromosome of each pair is the same, insert size was less than 250 bp, and strands were in opposite/inward orientation ($\Rightarrow \Leftarrow$). This was done using bedtools to convert the BAM files to Paired BED files (bedpe) and then filtering. The filtered read pairs were used to compute genome-wide coverage using the bedtools genomecov program.

Raw and processed reads have been deposited in the GEO database under accession number GSE71929. This accession also contains text files containing the calculated raw occupancy score at each base position in the genome.

Analysis of human hotspots

For the human data from (28), hotspot strength comparison was performed with a subset of 37,345 hotspots present in at least one of the three individuals (AA₁, AA₂, AB₁). Hotspots with strength of 0 were excluded, since these might be arising from technical reasons instead of representing actual variation in hotspot strength. This exclusion means that the extent of complete hotspot loss in individuals will be underestimated. The DNA recognition preference of the *PRDM9* B allele is not detectably different from the DNA recognition preference of the *PRDM9* A allele, and therefore hotspots found in these three individuals are considered as defined by the same *PRDM9* allele (28, 50)

Estimating divergence in terms of sexual generations

Since yeasts more commonly undergo asexual reproduction and undergo sexual reproduction less frequently (~1 sexual cycle every 1000 mitotic divisions, estimates by (51) based on *S. paradoxus* population studies), perhaps a more equitable comparison of evolutionary divergence between species that undergo obligate sexual reproduction (e.g., human, chimp, mouse) and species that undergo facultative sexual reproduction (e.g., yeasts), is to compare divergence in terms of sexual generations by dividing millions of years of evolution by sexual cycle length, especially when considering changes that are incorporated during sexual cycles. The following are the values used in our estimation.

Human vs. chimp comparison:

5 My divergence

Sexual cycle length 20 years (human), 10 years (chimp)

$5,000,000 \text{ years} / 20 \text{ years} = 250,000 \text{ generations (human)}$ since divergence with last common ancestor shared with chimps

$5,000,000 \text{ years} / 10 \text{ years} = 500,000 \text{ generations (chimp)}$ since divergence with last common ancestor shared with humans

Human vs. mouse comparison:

75 My divergence

Sexual cycle length 20 years (human), 2 months (0.17 years) (mouse)

$75,000,000 \text{ years} / 20 \text{ years} = 3,750,000 \text{ generations (human)}$ since divergence with last common ancestor shared with mice

75,000,000 years/0.17 years = 44,000,000 generations (mouse) since divergence with last common ancestor shared with humans

S. cerevisiae vs. *S. kudriavzevii* comparison:

15 My divergence

1 sexual cycle every 1000 mitotic divisions

Estimated length of mitotic divisions in wild habitat: 6 h (Mitotic divisions in the lab environment are approximately every 90 minutes, under controlled, optimal temperature and nutrient availability. In the natural environment, we envision fluctuation in temperature and climate, and nutrient availability to result in longer division times, or even periods of non-dividing state (52).)

Sexual cycle length:

1000 mitotic divisions x 6 h = 6000 h (250 days, or 0.68 years)

1 sexual cycle every 0.68 years

15,000,000 years/0.68 years = 22,000,000 sexual generations divergence between *S. cerevisiae* and *S. kudriavzevii*

Inbreeding predominates over outcrossing in wild populations of yeast, either via mating of spores from the same tetrad (automixis), or autodiploidization (mating type switching followed by mating of cells from the same haploid spore clone), with frequencies estimated at 0.94 and 0.05, respectively (51). The estimated frequency of outcrossing is 0.01 (51, 53). Automixis yields a complex genomic mix of heterozygosity and homozygosity dependent on centromere linkage, mating type locus linkage to its centromere, and degree of heterozygosity in the parental cell that underwent meiosis. Autodiploidization yields an essentially entirely homozygous diploid (except at the mating type locus). Because the effects of biased gene conversion are only relevant at heterozygous loci, a conservative calculation of sexual divergence in *S. cerevisiae* and *S. kudriavzevii* would be to consider only the frequency of outcrossing in our calculations:

0.001 (sexual cycles per mitotic division) x 0.01 (outcrosses per sexual cycle) = 1 outcrossed sexual cycle every 100,000 mitotic divisions

100,000 mitotic divisions x 6 h = 600,000 h (25,000 days, or 1 outcrossed meiosis every 68.5 years)

15,000,000 years/68.5 years = 220,000 sexual generations divergence between *S. cerevisiae* and *S. kudriavzevii*

If we instead use the estimate for *S. cerevisiae* of 1 outcrossing event every 50,000 mitotic divisions (54), the number of sexual generations since the last common ancestor of *S. cerevisiae* and *S. kudriavzevii* is 440,000. These numbers are the same order of magnitude as the divergence between humans and chimps (250,000 human generations, 500,000 chimp generations). Therefore, yeast species in this study have undergone ample cycles of meiosis in a heterozygous state to detect erosion of hotspot alleles if meiotic drive from biased gene conversion is not opposed by other selective constraints.

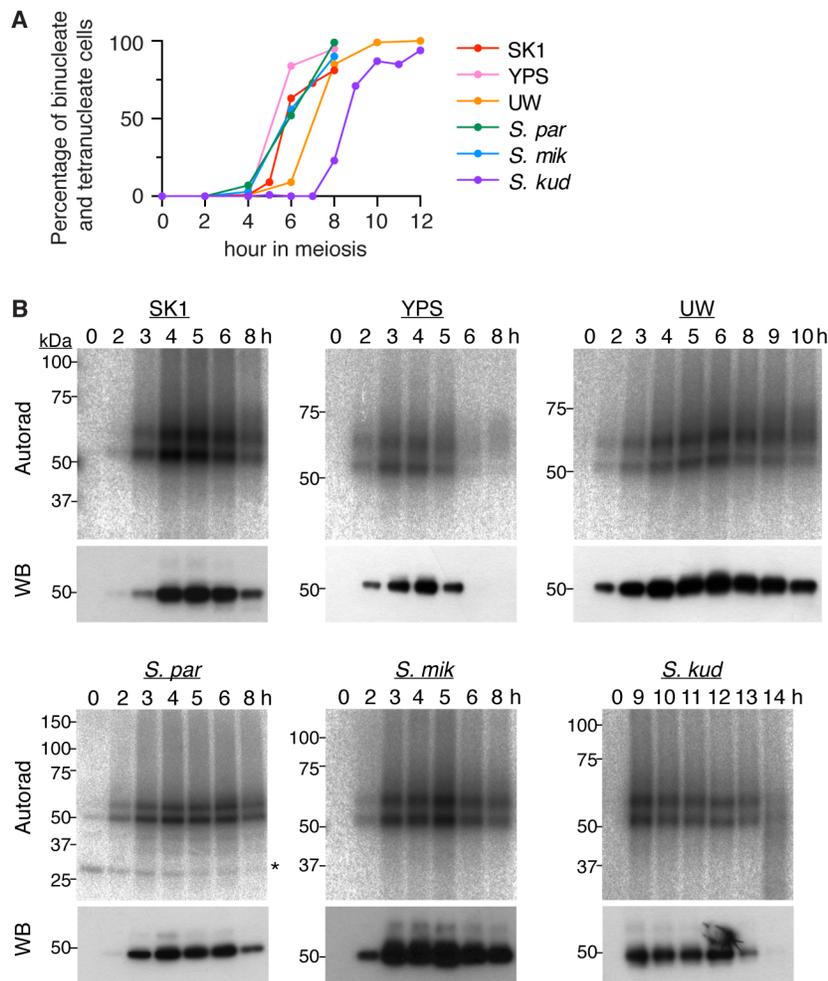


Fig. S1. Characteristics of sporulation and Spo11-oligo complexes in different *Saccharomyces* strains and species. (A) Meiotic progression showing percentage of cells completing the first division (total bi- and tetranucleate cells). ≥ 100 cells were counted at each time point for each sample. (B) Spo11-oligo complex time courses. Epitope-tagged Spo11 was immunoprecipitated with anti-Flag antibody from denaturing extracts of meiotic cultures at the indicated times, then radioactively labeled with terminal deoxynucleotidyl transferase and $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$. Labeling reactions were electrophoresed on SDS-PAGE gels. Radiolabeled Spo11-oligo complexes were detected by autoradiography (top) and total Spo11 was detected by anti-Flag western blot (WB). A substantial fraction of the Spo11-oligo complexes migrates as two prominent bands, reflecting two major size classes of Spo11 oligos (19). This pattern is conserved in all species examined here. Note that nearly all of the western blot signal is from free Spo11, i.e., protein that has not made a DSB (19). Asterisk indicates labeling of a nonspecific species in *S. paradoxus* that also appears when carrying out mock immunoprecipitation on an untagged Spo11 strain, but that is not visible when performing two sequential rounds of immunoprecipitation for Spo11-oligo mapping.

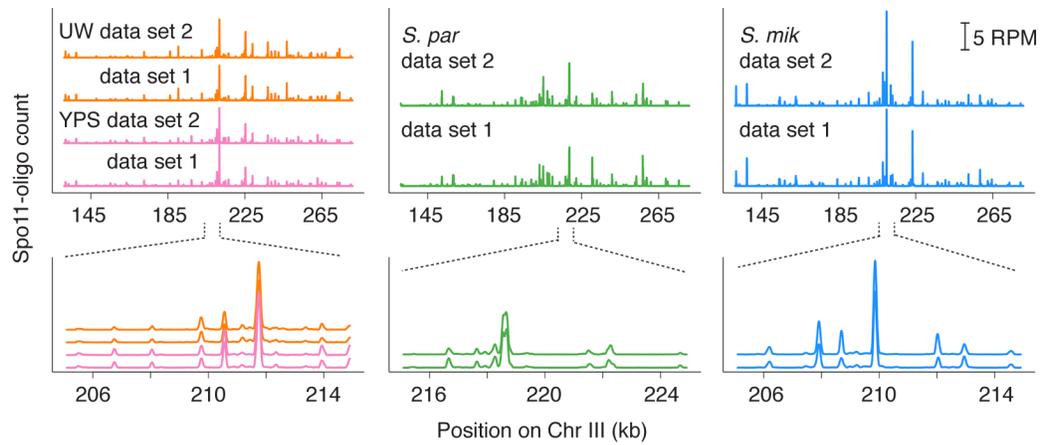


Fig. S2. Reproducibility of Spo11-oligo maps. Spo11-oligo counts were smoothed with a 201-bp Hann window.

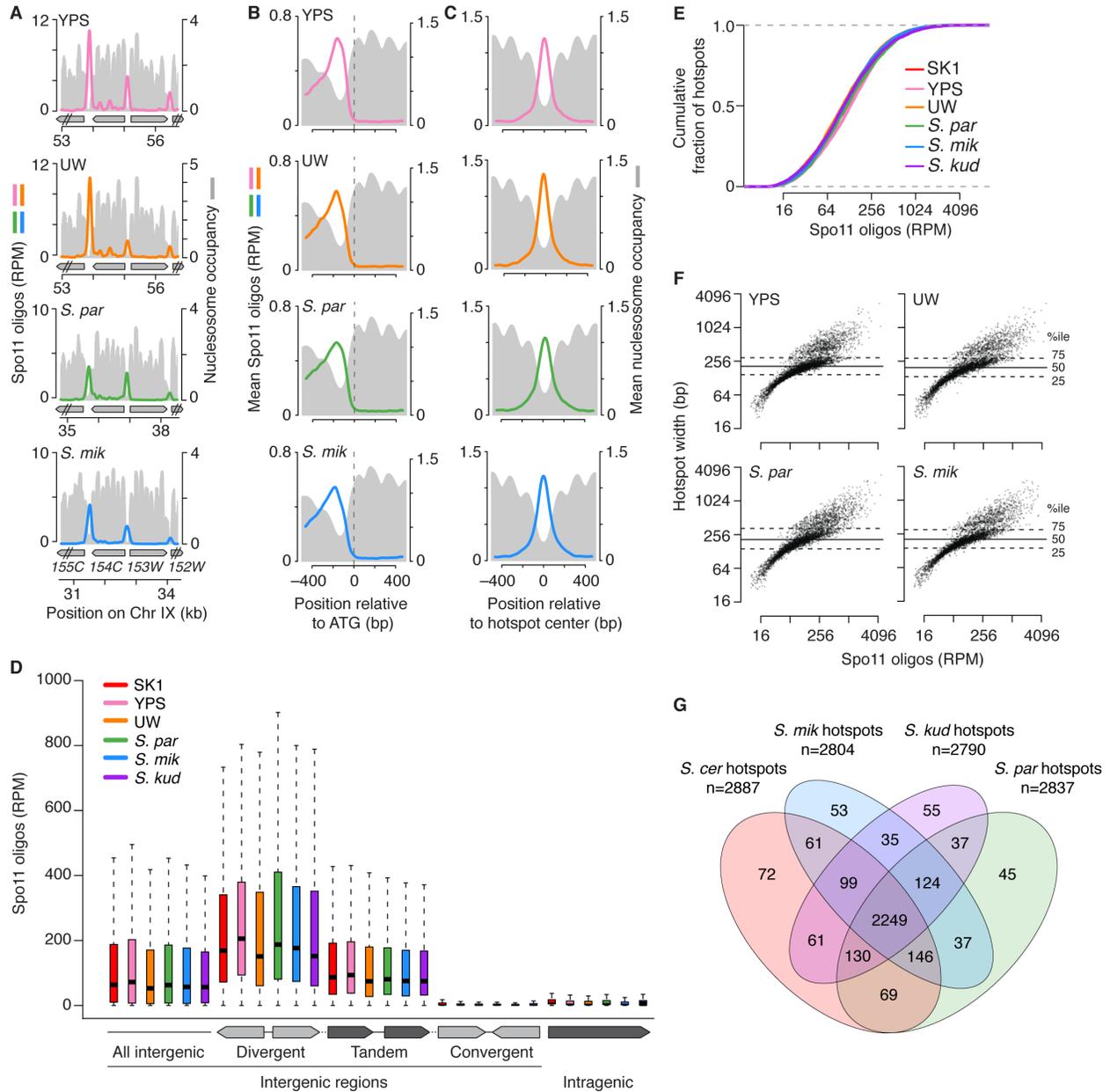


Fig. S3. Conserved targeting of DSBs to promoters. (A) Overlap of DSB hotspots with promoter NDRs. Data for additional *S. cerevisiae* strains and species not displayed in **Fig. 2A** are shown for the same sample region around *YIL154C*. Spo11-oligo maps were smoothed with a 201-bp Hann filter; nucleosome occupancy (MNase-seq read depth) was normalized to genome average (i.e., genome average = 1). (B) Average Spo11-oligo and nucleosome profiles around start codons, plotted as in **Fig. 2B** (*S. cer* strains, n=5766; *S. par*, n=5382 genes; *S. mik*, n=5684 genes). (C) Average Spo11-oligo and nucleosome profiles at hotspots, plotted as in **Fig. 2C** (*S. cer* YPS, n=4177; *S. cer* UW, n=3881; *S. par*, n=3833; *S. mik*, n=3829). (D) In all species examined, Spo11 oligos map preferentially to IGRs that contain promoters. Genomes were divided into genic and intergenic compartments, and IGRs were further subdivided according to the orientation of adjacent transcription units. Thick horizontal lines indicate medians, box edges

show the 25th and 75th percentiles, and whiskers indicate lowest and highest values within 1.5-fold of the interquartile range; outliers are not shown. The total number of IGRs and the breakdown by category in each species are as described in the Methods. The total number of genes are as follows: *S. cerevisiae*, 5766; *S. paradoxus*, 5382; *S. mikatae*, 5841; *S. kudriavzevii*, 5728. (E) Hotspot intensity varies over a similar smooth continuum in all strains/species. Cumulative plots of hotspots ranked by Spo11-oligo count (\log_2 scale) are shown (as in **Fig. 2E**). (F) Comparison of the distribution of hotspot widths vs. the distribution and Spo11-oligo counts for the strains and species not included in **Fig. 2F**. Data is plotted on a \log_2 scale but labeled according to a linear scale. (G) Conservation of promoter-associated hotspots. Using a set of 3426 stringently matched promoter-containing IGRs that could be unambiguously defined in all four species (see **Fig. 3A** and Methods), we first determined which hotspots called from each species' Spo11-oligo map overlapped such a promoter IGR. The Venn diagram in **Fig. 2G** shows how many called hotspots overlapped the same promoter IGR in both *S. cerevisiae* and *S. kudriavzevii*. Here, a four-way Venn diagram shows the overlap across all species examined. Most of these promoter-associated hotspots (2249) were shared between all four species, and another 499 were shared between three species. Promoter IGRs that were scored as hotspots in only one species were typically very weak, and one or more other species often yielded Spo11 oligos mapping to the same IGR but at levels below the arbitrary threshold set for the hotspot calling algorithm. These findings, along with the specific examples shown in **Fig. 2A**, **Fig. 3F,G**, and **fig. S3A**, demonstrate that hotspot positions in promoters are highly conserved across the *Saccharomyces sensu stricto* clade.

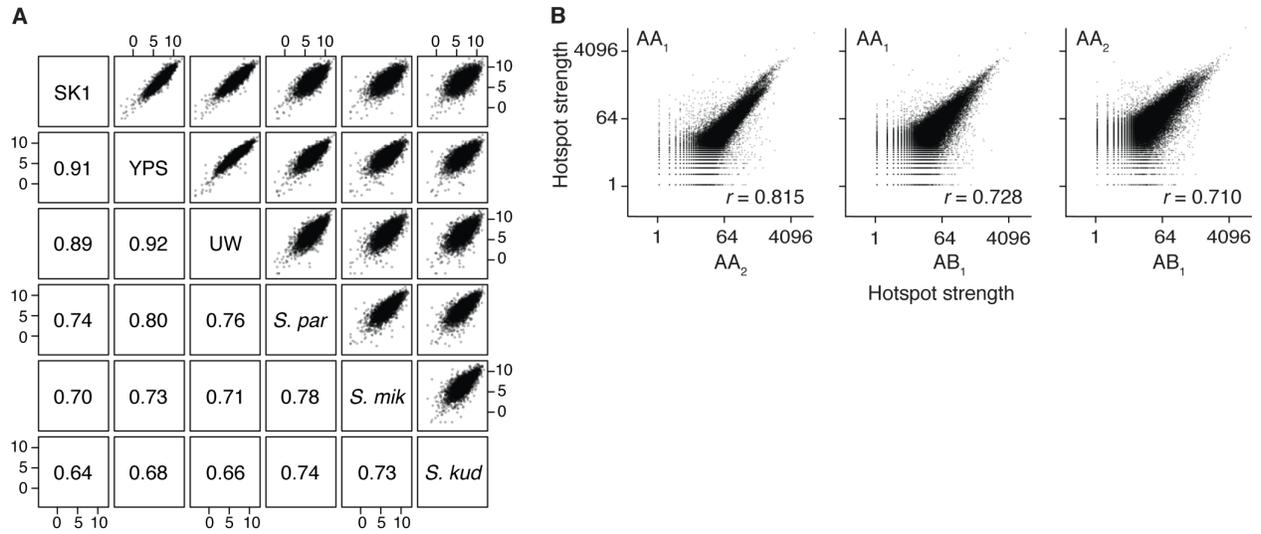


Fig. S4. Conservation of hotspot strength. (A) Comparison of Spo11-oligo counts (RPM, \log_2) within 3426 matched promoter IGRs among the different species/strains (expanded version of **Fig. 3B**). (B) Scatter plots comparing DSB hotspot strength between two men (designated AA₁ and AA₂) homozygous for the *PRDM9* A allele common in populations of European descent, and one man (designated AB₁) heterozygous for the A allele and the closely related B allele (data adapted from (28)). These *PRDM9* alleles recognize the same DNA sequence motif and thus define the same hotspot sites (28, 50). DSB activities were measured by deep-sequencing of single-stranded DNA co-immunoprecipitated with the DMC1 strand exchange protein (28, 55, 56).

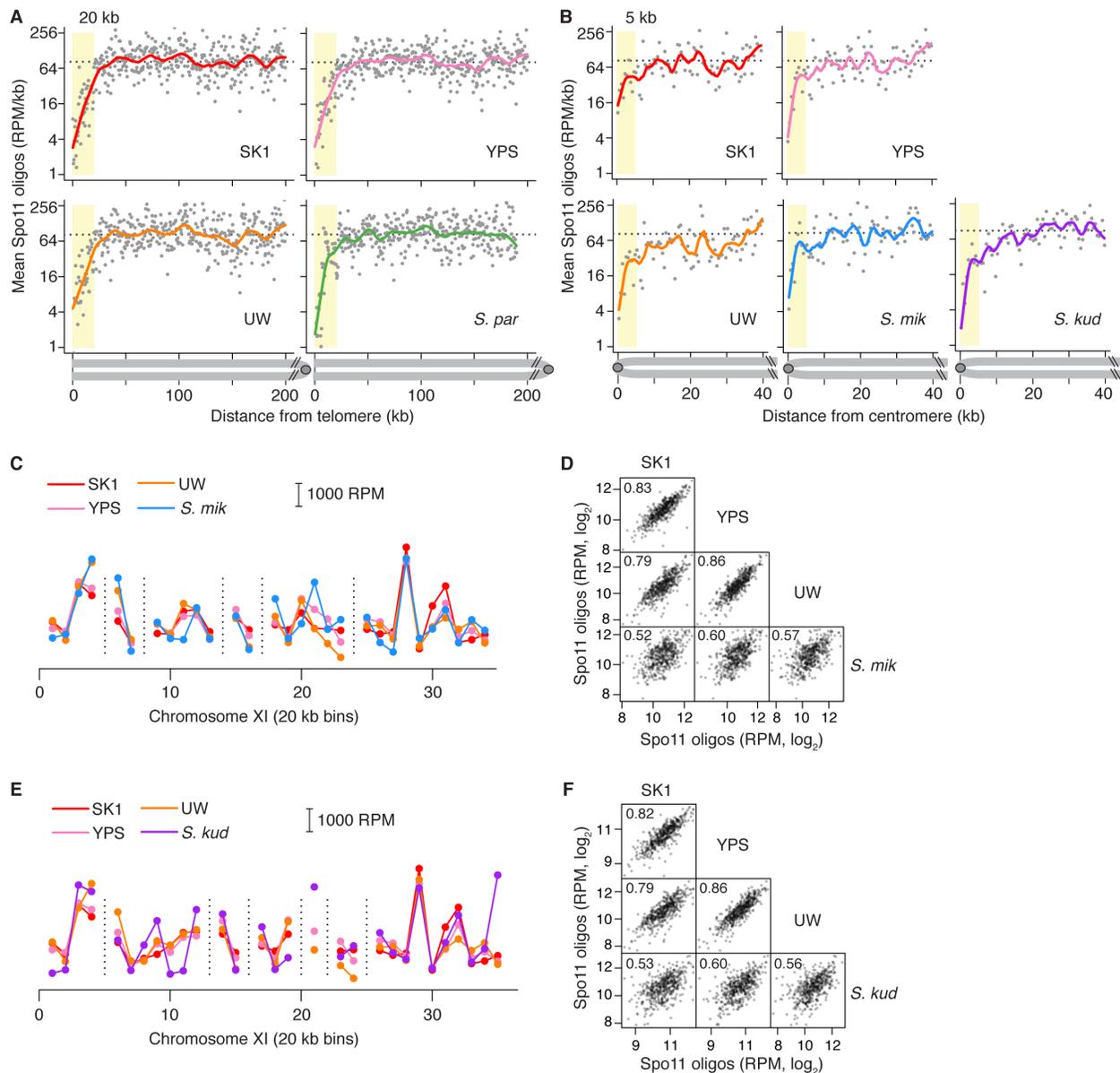


Fig. S5. Large-scale features of the DSB landscape are conserved. (A and B) Telomere-proximal and pericentric DSB suppression. Points are Spo11-oligo densities (plotted in \log_2 scale) in 500-bp bins averaged across all 32 chromosome arms. Dashed line indicates genome average in SK1; colored lines indicate smoothed fit (Lowess); yellow shading, DSB suppression zones. (C–F) Large-scale hot and cold interstitial domains are conserved. Interstitial segments (excluding 20 kb from chromosome ends and 10 kb from centromeres based on *S. cerevisiae* annotation) were defined as syntenic if orthologous genes were in the same order in pairwise comparisons of *S. cerevisiae* with *S. mikatae* (panels C,D) or with *S. kudriavzevii* (panels E,F). Spo11-oligo counts were then summed in these syntenic segments divided into 20-kb bins (**Table S5**). Panels C and E show the same representative genomic region as in **Fig. 4C**. Vertical dashed lines denote breaks in synteny. Panels D and F show genome-wide scatter plots and correlation coefficients as in **Fig. 4D**. Note that intra-species *S. cerevisiae* comparisons exhibit different correlation coefficients in the different figure panels because the correlations in a given panel are tested

within syntenic interstitial segments that are defined in one pairwise species comparison. Species pairs do not all share precisely the same blocks of synteny, so there are small differences as to which portions of the genome are being compared in each panel.

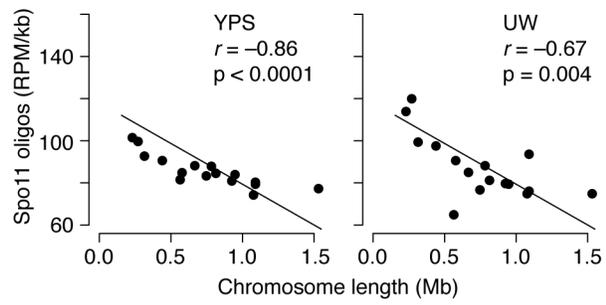


Fig. S6. Negative correlation between chromosome length and Spo11-oligo density is conserved in wild-derived *S. cerevisiae* strains.

Table S1. Yeast strains.

Strain number	Species	Genotype
SKY2522 ^a	<i>S. mikatae</i> IFO1815	<i>MATa/α</i>
SKY2523 ^a	<i>S. paradoxus</i> YPS138	<i>MATa/α</i>
SKY3821	<i>S. cerevisiae</i> SK1	<i>MATa/α; ura3"/; lys2"/; ho::LYS2"/; leu2Δ"/; arg4-bgl"/; nuc1Δ::LEU2"/; spo11-6His -3FLAG-loxP-kanMX-loxP"/;</i>
SKY4411	<i>S. paradoxus</i> YPS138	<i>MATa/α; spo11-6His-3FLAG-loxP-kanMX-loxP"/</i>
SKY4416 ^b	<i>S. kudriavzevii</i> ZP591	<i>MATa/α</i>
SKY4479 ^c	<i>S. cerevisiae</i> YPS128	<i>MATa/α</i>
SKY4481 ^c	<i>S. cerevisiae</i> UWOPS03-461.4	<i>MATa/α</i>
SKY4488	<i>S. kudriavzevii</i> ZP591	<i>MATa/α; spo11-6His-3FLAG-loxP-kanMX-loxP"/</i>
SKY4490	<i>S. mikatae</i> IFO1815	<i>MATa/α; spo11-6His-3FLAG-loxP-kanMX-loxP"/</i>
SKY4632	<i>S. cerevisiae</i> YPS128	<i>MATa/α; spo11-6His-3FLAG-loxP-kanMX-loxP"/</i>
SKY4664	<i>S. cerevisiae</i> UWOPS03-461.4	<i>MATa/α; spo11-6His-3FLAG-loxP-kanMX-loxP"/</i>

- a. From Ed Louis via Michael Lichten (National Cancer Institute)
b. FM1158 from Chris Todd Hittinger (Univ. Wisconsin-Madison)
c. From Ed Louis (Univ. Leicester)

Table S2. Mapping statistics for Spo11 oligo sequences.

Dataset	Strain	No. of reads (total reads)	Genome mapped to	No. mapped	No. mapped uniquely to genome
YPS128-3	<i>S. cerevisiae</i> YPS128	20,987,066	<i>S. cerevisiae</i> S288C	19,316,922	19,085,392 (98.8%)
YPS128-4	<i>S. cerevisiae</i> YPS128	24,963,684	<i>S. cerevisiae</i> S288C	22,743,707	22,471,981 (98.8%)
UWOPS-1	<i>S. cerevisiae</i> UWOPS03-461.4	5,871,998	<i>S. cerevisiae</i> S288C	5,087,159	4,996,549 (98.2%)
UWOPS-2	<i>S. cerevisiae</i> UWOPS03-461.4	3,799,030	<i>S. cerevisiae</i> S288C	3,416,873	3,358,908 (98.3%)
Spar1	<i>S. paradoxus</i> YPS138	6,641,178	<i>S. paradoxus</i> YPS138	6,005,923	5,926,000 (98.7%)
Spar2	<i>S. paradoxus</i> YPS138	6,872,508	<i>S. paradoxus</i> YPS138	5,675,981	5,602,019 (98.7%)
Spar1_CBS432 ^a	<i>S. paradoxus</i> YPS138	6,641,178	<i>S. paradoxus</i> CBS432	4,625,865	4,534,439 (98.0%)
Spar2_CBS432 ^b	<i>S. paradoxus</i> YPS138	6,872,508	<i>S. paradoxus</i> CBS432	4,322,128	4,244,734 (98.2%)
Smik2	<i>S. mikatae</i> IFO1815	3,112,231	<i>S. mikatae</i> IFO1815	2,937,039	2,916,164 (99.3%)
Smik3	<i>S. mikatae</i> IFO1815	6,665,306	<i>S. mikatae</i> IFO1815	6,297,240	6,241,580 (99.1%)
Skud2	<i>S. kudriavzevii</i> ZP591	5,016,979	<i>S. kudriavzevii</i> ZP591	4,597,476	4,558,486 (99.2%)
Skud3	<i>S. kudriavzevii</i> ZP591	5,778,620	<i>S. kudriavzevii</i> ZP591	5,358,181	5,282,452 (98.6%)

a. Same sample as Spar1, but reads were mapped to *S. paradoxus* type strain CBS432

b. Same sample as Spar2, but reads were mapped to *S. paradoxus* type strain CBS432

The following are provided as separate Excel files:

Table S3. Hotspot lists in *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*.

Table S4. List of 3426 matched promoter intergenic regions.

Table S5. List of 20-kb bins for large-scale interstitial analysis.