File S1

Materials and Methods

Strain Construction: The genotypes of all strains in this study are given in Table S1. In all of the diploids used in our mapping studies, one of the haploid parental strains was derived by transformation of the haploid J178-1d (Brock and Bloom, 1994). J178-1d was created by a complex series of crosses, at least one of which involved the haploid strain S288c. In one derivative of J178-1d (J178#7-20), a conditional centromere (*GAL-CEN3*) and a *URA3* gene were inserted within the *HIS4* locus (SGD coordinates 65934-68333) of chromosome III (*his4::GAL CEN3::URA3*) (Brock and Bloom, 1994). The orientation of the conditional centromere is the same as that of the wild-type *CEN3*.

In another haploid derivative of J178-1d (MG42), the conditional centromere and *URA3* gene replaced the *CAN1* locus (SGD coordinates 31694-33466) on the left arm of chromosome V (*can1-Δ::GAL-CEN3::URA3*). This strain was constructed by transforming J178-1d using a PCR fragment obtained by amplifying plasmid pR285 #7/pR285-GALCEN3#7 DNA (Brock and Bloom, 1994) with the primers GALCEN3/CAN F and GALCEN3/CAN R; the sequences of all primers used in strain constructions are in Table S2. The resulting PCR fragment contains the *GAL-CEN3 URA3* cassette with sequences derived from the *CAN1* locus at the ends of the fragment. We selected Ura^+ transformants and screened those transformants for resistance to canavanine. We also confirmed the location of cassette by PCR using primer pairs CANupF with URA3R, and CANdnR with pBR322Ftest. We confirmed that the *CAN1* gene was deleted by using the primer pair CANF and CANR. The orientation of the conditional centromere inserted on chromosome V is opposite to that of the conditional centromere inserted on chromosome III in J178#7-20.

In the haploid strain MG48, the conditional centromere and the *URA3* marker were located on the left arm of chromosome V replacing the region between Saccharomyces Genome Database (SGD) coordinates 80163 and 80362 (*V80163-80362Δ::URA3::GAL-CEN3*). This strain was constructed by transforming J178-1d using a PCR fragment obtained by amplifying plasmid pR285 #7/pR285-GALCEN3#7 DNA with the primers GALCEN3/80k F and GALCEN3/80k R, and selecting Ura^+ transformants. The insertion of the conditional centromere at the correct site was confirmed using three sets of primer pairs: 80kupF with URA3R, 80kdnR with pBR322Ftest, and 80kupF with 80kdnR.

We also used a haploid strain (PSL5) derived from the sequenced clinical isolate YJM789 (Wei *et al.*, 2007); PSL5 (*MATα ade2-1 ura3 can1Δ::SUP4-o gal2 ho::hisG*) has been described previously (Lee *et al.*, 2009). The diploids used in our study were constructed by the following crosses: WS49 (J178-#7-20 x PSL5); WS83 (MG42 x PSL5); WS92 (MG48 x PSL5). All strains are

heterozygous for SNPs located throughout the genome (St. Charles *et al.*, 2012). WS49, WS83, and WS92 are isogenic except for the location of the conditional centromere.

Analysis of structural differences between the chromosome III homologs derived from J178-1d and YJM789

In addition to SNPs that distinguish the chromosome III homologs derived from J178-#7-20 and PSL5, there were also insertions of Ty elements and other structural changes. It is have been noted previously (Hill and Bloom, 1989; Wicksteed *et al.*, 1994) that chromosome III from J178-1d is larger than observed in most other yeast strains. Using ORF-containing microarrays and procedures described in McCulley *et al.* (2010), we found that J178-#7-20 had a duplication on the right arm of chromosome III that included ORFs between *YCR019W* and *YCR027C*. The location of this duplication suggests that it was generated by unequal crossing-over between two previously-mapped pairs of Ty elements termed FS1 and FS2 (Umezu *et al.*, 2002; Lemoine *et al.*, 2005).

We also examined Ty elements located on the left arm of chromosome III in the two strains by Southern analysis and a series of PCR reactions. The S288c-related strain described in SGD has a single Ty2 element located between *KCC4* and *LEU2*, although some other strains contain both Ty1 and Ty2 between these two genes. We isolated genomic DNA from MS71 (a control wild-type strain; Sia *et al.*, 1997), PSL5, and J178-#7-20, and treated the samples with *Acl*I, a restriction enzyme that does not cut within Ty elements. The resulting fragments were examined by standard Southern analysis, using a probe containing *KCC4* sequences; this probe was prepared using genomic DNA and primers RKCC4-1 and KCC4-1 (Table S2). The observed sizes of the *Acl*I restriction fragment hybridizing to the probe were 6.5 kb, 11 kb, and 18 kb for the strains PSL5, MS71, and J178-#7-20, respectively. These sizes suggest that PSL5 has no Ty element near *LEU2*, MS71 has one Ty element, and J178-#7-20 has two Ty elements. This conclusion was confirmed by analyzing fragments generated by double digests of genomic DNA with *Nco*I and *Nsi*I. The observed sizes of fragments hybridizing to the *KCC4* probe were 8 kb (PSL5), 5 kb (J178-#7-20), and 14 kb (MS71). Since *Nsi*I has three recognition sites in Ty1, but not in Ty2, these results support the conclusion that PSL5 lacks Ty elements near *LEU2*, MS71 has one Ty2 element, and J178-#7-20 has closely-linked Ty1 and Ty2 elements.

These conclusions were further supported by PCR analysis. Using primers KCC4F2 82211 and Ty2R 85162 (a primer with homology to both Ty1 and Ty2), we observed no amplification with PSL5 genomic DNA, an 800 bp fragment with J178-#7-20 DNA, and a 3 kb fragment with MS71 DNA. Using primers KCC4F2 82211 and Ty2R 85553 (a primer that is Ty2-specific), we observed a low level of a fragment of 1.6 kb with PSL5, a 10 kb fragment with J178-#7-20 DNA, and a 3.3 kb fragment with MS71 DNA. Assuming that the low level of the 1.6 kb fragment observed with PSL5 is non-specific, these results argue that PSL5 lacks a Ty2 element, J178-#7-20 has a centromere-distal Ty1 element and a centromere-proximal Ty2 element, and MS71 has only a Ty2 element. We confirmed that J178-#7-20 has a Ty1 element in the Watson orientation using two PCR primer pairs:

Ty1F with KCC4 F2 82211, and Ty1R with KCC4 F2 82211. Only the PCR reaction of Ty1R and KCC4 F2 82211 yielded the 800 bp fragment as expected if the Ty1 element was in Watson orientation.

Identification of strains with recombination events induced by dicentric chromosome breakage

The *GAL-CEN3* conditional centromere is inactive in cells grown in medium containing galactose and active in cells grown in glucose (Hill and Bloom, 1987). In our experiments, all diploid strains were grown from single cells to colonies on solid medium containing galactose (YPGal) at 30°C for two days. Individual colonies were selected from these plates, and re-streaked on plates containing glucose (YPD) and incubated at 30°C for two days. The resulting colonies were then replica-plated to YPGal medium lacking uracil to identify derivatives that had lost the *URA3* marker adjacent to the conditional centromere. The percentages of colonies that were either Ura^- or sectored Ura^+/Ura^- in cells grown on YPD-containing plates (median value of five independent cultures) were 78% (WS49), 94% (WS83), and 93% (WS92). The median percentages of Ura^- or sectored Ura^+/Ura^- colonies (median value of five independent cultures) in cells grown on galactose-containing plates (inactive conditional centromere) and then plated on galactose-containing plates were 4% (WS49), 2% (WS83), and 2% (WS92).

Analysis of loss of heterozygosity (LOH) using restriction digests of PCR fragments

In yeast strains that are heterozygous for markers, mitotic crossovers can generate loss of heterozygosity of markers centromere-distal to the crossover (Lee *et al.*, 2009; St. Charles *et al.*, 2012). The transition between heterozygous markers and homozygous markers, therefore, locates the position of the crossover. We looked for LOH using two procedures, an approach in which LOH was detected by a PCR-based approach (describe below), and an approach utilizing oligonucleotide-containing microarrays. For the first approach, we used genomic DNA sequence information to identify SNPs that distinguished the two haploid strains in the region between the conditional centromere and the natural centromere on chromosome III. We then determined which of these SNPs altered a restriction site, and designed primers that would amplify a region of several hundred bp flanking the SNP. For example, there is a SNP at SGD coordinate 70426 that results in an *Hpy*166II site in the PSL5 strain that is absent in the J178-#7-20 strain. We designed primers flanking this site that produce a fragment of about 500 bp. If we treat this fragment produced by PCR amplification from the heterozygous diploid strain with *Hpy*166II and analyze the fragments by gel electrophoresis, we obtain three fragments: about 500 bp (representing the SNP derived from the J178-#7-20 strain), and about 360 and 140 bp (representing the SNP derived from the PSL5 strain). In diploids that undergo LOH for a SNP at this position, we observe either one 500 bp fragment or two fragments of 360 and 140 bp. The location of the SNPs, the primers used to produce the restriction fragments, and the diagnostic restriction enzyme are shown in Table S3.

In the WS49 strain, which has the conditional centromere on chromosome III, we first examined polymorphic markers located centromere-distal to the conditional centromere (33 and 58), within the intercentromeric region (68, 81, 102, 103, 111, 113), and located on the opposite chromosome arm (116) for LOH; the names of the SNPs reflect their approximate SGD coordinates in kb (Table S3). As explained in the Main Text, based on the patterns of LOH observed with these markers, we classified the Ura$^-$ derivatives of WS49 as Class 1, 2, 3, or 4. We then performed more detailed mapping of recombination breakpoints in Class 1 strains using other markers (70, 74, 78, 91, 96, 98, 106, and 107). The conditional centromere on III is inserted near SGD coordinate 67 kb, and the natural centromere is near coordinate 114 kb.

By a similar approach, we mapped crossovers in strains with the conditional centromere on chromosome V. For these experiments, we used the primers and restriction enzymes described in Lee *et al.* (2009). For the preliminary mapping of strain WS83, we used the chromosome V markers 7, 25, 41, 70, 112, 133, and 561; the conditional centromere is inserted near SGD coordinate 32 kb, and *CEN5* is located at SGD coordinate 152 kb. The other markers used for mapping are: 35, 43, 44, 46, 49, 52, 55, 56, 57, 60, 64, 76, 80, 83, 87, 92, 94, 99, 104, 108, 114, 115, 117, 119, 122, 126, 141, 144, 147, and 151. For the preliminary mapping of strain WS92, we used the chromosome V markers 7, 52, 76, 83, and 152. The other markers used for mapping are: 87, 92, 94, 99, 104, 108, 112, 119, 133, 141, 144, 147, and 151.

The primers and restriction enzymes used to analyze heterozygous SNPs on chromosome V are in Table S2 of Lee *et al.* (2009) with the exception of markers 7, 25, 152, and 561. Markers 7 and 25 are located centromere-distal to the conditional centromeres which are located at 32 kb in MS83 and 80 kb in WS92. Marker 7 is located at SGD coordinate 7005. For this polymorphism, we amplified genomic DNA with the forward primer ATCCTCATCTTACCAGCTCACTC (starting coordinate at 6857) and reverse primer AGTAGTACCTGTTTTAATGGG (starting coordinate at 7249). The diagnostic restriction enzyme was *Dra*I, which cuts the genomic DNA of MG42 and MG48, but not that of PSL5; both MG42 and MG48 are derived from J178-1d. The PCR fragment used to check marker 25 was generated with the primers 5' CACTTGAGGCCACGCATACTG and 5' GCAACGTTTGGGAAGAAAACG. We tested Ura$^-$ strains derived from WS83 and WS92 for LOH of a telomere-associated SNP using a polymorphism located on chromosome V at SGD coordinate 24903 (marker 25). The YJM789-derived homolog has a *Hind*III site at this position that is absent in J178-1d derivatives. The primers used to generate the PCR fragment with this polymorphism were 5' CACTTGAGGCCACGCATACTG and 5' GCAACGTTTGGGAAGAAAACG.

For strains WS83 and WS92, we used different markers located on the chromosome V arm opposite the conditional centromere. For WS83, we used a marker (561) located at 560715. For this polymorphism, we amplified genomic DNA with the

forward primer TTCTCAGCCGTACAATCATGC (starting coordinate at 560490) and reverse primer AAACTCCTTCCAAAGGGTCTGG (starting coordinate at 560980). The diagnostic restriction enzyme was *Eco*RI, which cuts the genomic DNA of PSL5, but not that of MG42. For WS92, we used a marker (152) located at 152163. For this polymorphism, we amplified genomic DNA with the forward primer TTGGTAAACAAAGGGCCAAGC (starting coordinate at 151849) and reverse primer ATGTGCGGCTTTGTCAGCAG (starting coordinate at 152295). The diagnostic restriction enzyme was *Cac*8I, which cuts the genomic DNA of MG48, but not that of PSL5.

Analysis of loss of heterozygosity (LOH) using SNP microarrays

Three recombination events derived from WS49 and four events derived from WS83 were mapped by both the PCR/restriction enzyme method described above and by oligonucleotide-containing microarrays. Because the stability of short duplexes is sensitive to mismatches, it is possible to design an oligonucleotide-containing microarray that can distinguish whether SNPs are heterozygous or homozygous (Gresham *et al.*, 2010). Previously, we designed Agilent SNP arrays to look for LOH throughout the genome in a diploid formed by a cross of W303a (closely related to S288c) and YJM789 (St. Charles *et al.*, 2012). For each SNP analyzed (about 15,000 distributed throughout the genome), four 25-base oligonucleotides were used, two identical to the Watson and Crick strands of the W303a-specific SNP and two identical to the Watson and Crick strands of the YJM789-specific SNP; the polymorphism was the central base of the oligonucleotide. The sequences of the oligonucleotides used in the construction of the microarray are given in St. Charles *et al.* (2012).

To look for LOH in strains with a recombination event, we labeled genomic DNA from the control strain with a Cy5-tagged nucleotide and DNA from a control heterozygous strain with Cy3-tagged nucleotide. The labeled samples were mixed and hybridized to the microarray. Following hybridization, the arrays were scanned and the ratio of hybridization to the two samples was determined as described previously (St. Charles *et al.*, 2012). Ratios of hybridization for each oligonucleotide were normalized to the Cy5/Cy3 ratio of all of the oligonucleotides on the microarray. In general, the recombination breakpoints determined by the PCR-based method were in good agreement with those determined by microarrays (discussed further in Results).

Statistical Analysis

We performed two types of statistical tests. First, we determined whether the distributions of chromosome break sites in the diploid strain were significantly different from a random distribution. For this analysis, we divided the region between the conditional centromere and the natural centromere into approximately equal-sized intervals. Based on the number of recombination events mapped for each diploid, we calculated the expected number of events in each interval and these numbers were compared to the observed distribution by chi-square analyses; these chi-square tests were done the VassarStats Website (http://faculty.vassar.edu/lowry/VassarStats.html). For WS49, the intervals used (markers shown in parentheses) were: Interval 1 (68 to 81); Interval 2 (81-91), Interval 3 (91-102), and Interval 4 (102-115). For WS83, the physical intervals were: Interval 1 (33-49), Interval 2 (49-64), Interval 3 (64-80), Interval 4 (80-94), Interval 5 (94-112), Interval 6 (112-126), Interval 7 (126-141), and Interval 8 (141-152). For the diploid WS92, we used Intervals 4 to 8 as specified for WS83.

As will be described in the Results section, for all three diploids, we found that the region located approximately 10 kb centromere-proximal to the conditional centromere had an elevated frequency of recombination events. Regions located outside of this "hotspot" were examined to determine if various chromosome elements (replication origins, palindromic sequences, and other elements described below) were over-represented at the recombination breakpoints. For purposes of this calculation, we excluded the intercentromeric region located near the conditional centromere. The lengths of the mapped regions (MRs) examined and the number of events for each strain were WS49 (33538 kb, 14), WS83 (103088 kb, 15), and WS92 (57657 kb, 8). These MRs were calculated by subtracting the SGD coordinate corresponding to the most centromere-distal SNP used in the mapping (excluding the hotspot) from the SGD coordinate representing the boundary of the "natural" centromere. The sum of the MRs for each strain is equivalent to the MR multiplied by the number of events. Thus, the sums of the MRs are 469532 kb (WS49), 1546320 kb (WS83), and 461256 kb (WS92). We define the recombination breakpoints (RBs) as the distance separating the last heterozygous site from the first homozygous site. The analysis was performed in several steps. First, for each strain, we summed the lengths of the RBs over all of the mapped events within the MR; the sum of the RBs for each strain were 72651 kb (WS49), 68900 kb (WS83), and 35469 kb (WS92). Second, we determined the lengths of sequences that are not RBs (NRBs) for all of the mapped events (equivalent to the sum of the RBs subtracted from the sum of the MRs for each strain); the NRBs were 396881 kb (WS49), 1477420 kb (WS83), and 425787 kb (WS92). Third, for each element, we calculated its density within the MR (the number of elements divided by the MR distance). Fourth, we calculated the expected number of elements within the summed RBs by multiplying the density of the chromosomal element by the summed RBs; we calculated the

expected number of elements in the NRBs in a similar way. We then compared the expected numbers of elements within and outside of the RBs with the observed numbers within and outside of the RBs by chi-square analysis.

To illustrate the method, we will describe our analysis to determine whether palindromic sequences are over-represented in recombination breakpoints (RBs) in the WS49 data. To determine the location of the palindromic sequences ($\geq$ 16 base pairs), we used the data in Lisnic *et al.* (2005). There are four palindromic sequences located in the 33538 kb MR between SGD coordinates 80845 and 114383, a density of 0.00012/kb. Since we had 14 events, the sum of palindromes for all the events is 56. In WS49, we observed that five palindromes were in the RB regions and 51 palindromes were in the NRB regions. The expected number of palindromes in the RB regions, assuming a random distribution, is 0.00012/kb x 72651 kb or 8.7; the expected number in the NRB regions is 0.00012 x 396881 or 47.3. When the observed and expected numbers were compared by chi-square test, the p value was 0.24, indicating that palindromes are not significantly enriched at the recombination breakpoints.

A similar analysis was done for each strain with the following elements (descriptions of the element and references in parentheses): tandem repeats (repeats between 2 and 213 bp with a minimum repeat tract of 24 bp; Gelfand *et al.*, 2007), G4 DNA (four tracts of 3 G's separated by spacers <25 bp; Capra *et al.*, 2010), tRNA genes (SGD), ARS elements (SGD), triplet repeats ($\geq$ 8 repeats; Gelfand *et al.*, 2007), long terminal repeats (SGD), peaks of gamma-H2AX (Szilard *et al.*, 2010), Rrm3p pause sites (Azvolinsky *et al.*, 2009), and replication-termination regions (Fachinetti *et al.*, 2010). We also looked for correlations with highly-transcribed genes. For this analysis, we determined the number of highly-transcribed genes in the MRs for all three strains, defining a highly-transcribed gene as a gene ranking in the top 20%, using the database of Nagalakshmi *et al.* (2008). For each of these genes, we calculated the midpoint of the transcript, and then determined whether these midpoints were over-represented in the RBs.

For the three individual strains, none of the examined elements had a significant over-representation at the breakpoints when corrections were performed for multiple comparisons. Since the number of events per strain was small, we also examined each element summed over all three strains. For this analysis, we added the numbers of observed and expected events for each category of element, and performed a chi-square analysis on the totals. None of the elements was significantly (p value <0.05 after correction for multiple comparisons) over-represented in this analysis.

Literature cited

Azvolinsky, A., P. G. Giresi, J. D. Lieb, and V. A. Zakian, 2009 Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. Mol. Cell 34: 722-734.

Brock, J. A., and K. Bloom, 1994 A chromosome breakage assay to monitor mitotic forces in budding yeast. J. Cell Sci. 107: 891-902.

Capra, J. A., K. Paeschke, M. Singh, and V. A. Zakian, 2010 G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genome features in *Saccharomyces cerevisiae*. PLoS Comp. Biol. 6: e1000861.

Fachinetti, D., R. Bermejo, A. Cocito, S. Minardi, Y. Katou *et al.*, 2010 Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. Mol. Cell 39: 595-605.

Gelfand, Y., A. Rodriguez, and G. Benson, 2007 TRDB—the Tandem Repeats Database. Nucleic Acids Res. 35: D80-87.

Gresham, D., B. Curry, A. Ward, D. B. Gordon, L. Brizuela *et al.*, 2010 Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. Proc. Natl. Acad. Sci. U.S.A. 107: 1482-1487.

Hill, A., and K. Bloom, 1987 Genetic manipulation of centromere function. Mol. Cell. Biol. 7: 2397-2405.

Hill, A., and K. Bloom, 1989 Acquisition and processing of a conditional dicentric chromosome in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 9: 1368-1370.

Lee, P. S., P. W. Greenwell, M. Dominska, M. Gawel, M. Hamilton *et al.*, 2009 A fine-structure map of spontaneous mitotic crossovers in the yeast *Saccharomyces cerevisae*. PLoS Genet. 5: e1000410.

Lemoine, F. J., N. P. Degtyareva, K. Lobachev, and T. D. Petes, 2005 Chromosome translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites. Cell 120: 587-598.

Lisnic, B., I. K. Svetec, H. Saric, I Nikolic, and Z. Zgaga, 2005 Palindrome content of the yeast *Saccharomyces cerevisiae* genome. Curr. Genet. 47: 289-297.

McCulley, J. L., and T. D. Petes, 2010 Chromosome rearrangements and aneuploidy in yeast strains lacking both Tel1p and Mec1p reflect deficiencies in two different mechanisms. Proc. Natl. Acad. Sci. U.S.A. 107: 11465-11470.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha *et al.*, 2008 The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320: 1344-1349.

Narayanan, V., P. A. Mieczkowski, H.-M. Kim, T. D. Petes, and K. S. Lobachev, 2006 The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. Cell 125: 1283-1296.

Sia, E. A., R. J. Kokoska, M. Dominska, P. Greenwell, and T. D. Petes, 1997 Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol. Cell. Biol. 17: 2851-2858.

St. Charles, J., E. Hazkani-Covo, Y. Yin, S. L. Andersen, F. S. Dietrich *et al.*, 2012 High-resolution genome-wide analysis of irradiated (UV and gamma-rays) diploid yeast cells reveals a high frequency of genomic loss of heterozygosity (LOH) events. Genetics 190: 1267-1284.

Szilard, R. K., P. E. Jacques, L. Laramee, B. Cheng, S. Galicia *et al.*, 2010 Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. Nature Struct. & Mol. Biol. 17: 299-305.

Umezu, K., M. Hiraoka, M. Mori, and H. Maki, 2002 Structural analysis of aberrant chromosomes that occur spontaneously in diploid *Saccharomyces cerevisiae*: retrotransposon Ty1 plays a crucial role in chromosomal rearrangements. Genetics 160: 97-110.

Wei, W., J. H. McCusker, R. W. Hyman, T. Jones, Y. Ning *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. Proc. Natl. Acad. Sci. U.S.A. 104: 12825-12830.

Wicksteed, B. L., I. Collins, A. Dershowitz, L. I. Stateva, R. P. Green *et al.*, 1994 A physical comparison of chromosome III in six strains of *Saccharomyces cerevisiae*. Yeast 10: 39-57.