Developmental Cell, volume 15

Supplemental Data

Global Analysis of the Meiotic Crossover Landscape

Stacy Y. Chen, Tomomi Tsubouchi, Beth Rockmill, Jay S. Sandler, Daniel R. Richards, Gerben Vader, Andreas Hochwagen, G. Shirleen Roeder, and Jennifer C. Fung

Supplemental Results

Minor Alterations of Crossing Over in the Hybrid Background

Studies have shown that in homologous strains (10-30% sequence divergence), spore viability is reduced to less than 1.0% and recombination to less than 10% of normal levels (Hunter et al., 1996), calling into question whether CO control is "normal" in strains showing any amount of divergence. However, this is not a concern for the S96/YJM789 hybrid used in this study, since spore viability is relatively high (81%, n=541). We also measured recombination and interference genetically in a hybrid between one parent (YJM789) and a lab strain (BR1919-8B (also an S288 derivative)) and compared it to the homozygote (BR1919 2n). Note that the BR1919-8B strain, rather than S96, was chosen for comparison purposes since the BR1919-8B strain was already well-characterized for both recombination and interference. Although there are differences in the CO levels in the two intervals examined (typical for strains heterozygous at multiple loci), no significant difference in interference levels were found (Figure S1A, Figure S1B).

Supplemental Figures and Tables

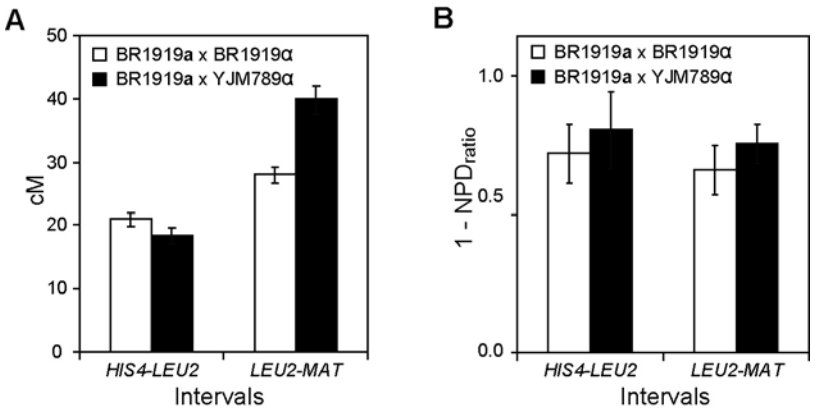

Figure S1. Characterization of Hybrid Strain

(A) Comparison of crossing over for two intervals on chromosome III measured in map distances for the BR1919-8B diploid strain and for a diploid resulting from mating BR1919-8Ba to YJM789. Map distances are significantly different (difference in map distances were greater than twice the SE) for the *LEU2-MAT* (LM) interval ($2*SE < |LM_1-LM_2|$: $0.0518 < 0.1192$), but not for *HIS4-LEU2* (HL) ($2*SE < |HL_1-HL_2|$: $0.0336 < 0.0265$). (B) Comparison of interference (1-NPD ratio) between the same strains for the same intervals as in (A) measured by tetrad dissection (n = 1000, BR1919-8B diploid; n = 515, BR1919-8Ba x YJM789). Chi-square tests show no difference in interference between the two strains (*HIS4-LEU2*: $\chi^2=0.0014$, $P > 0.95$; *LEU2-MAT*: $\chi^2=0.0305$, $P > 0.5$).

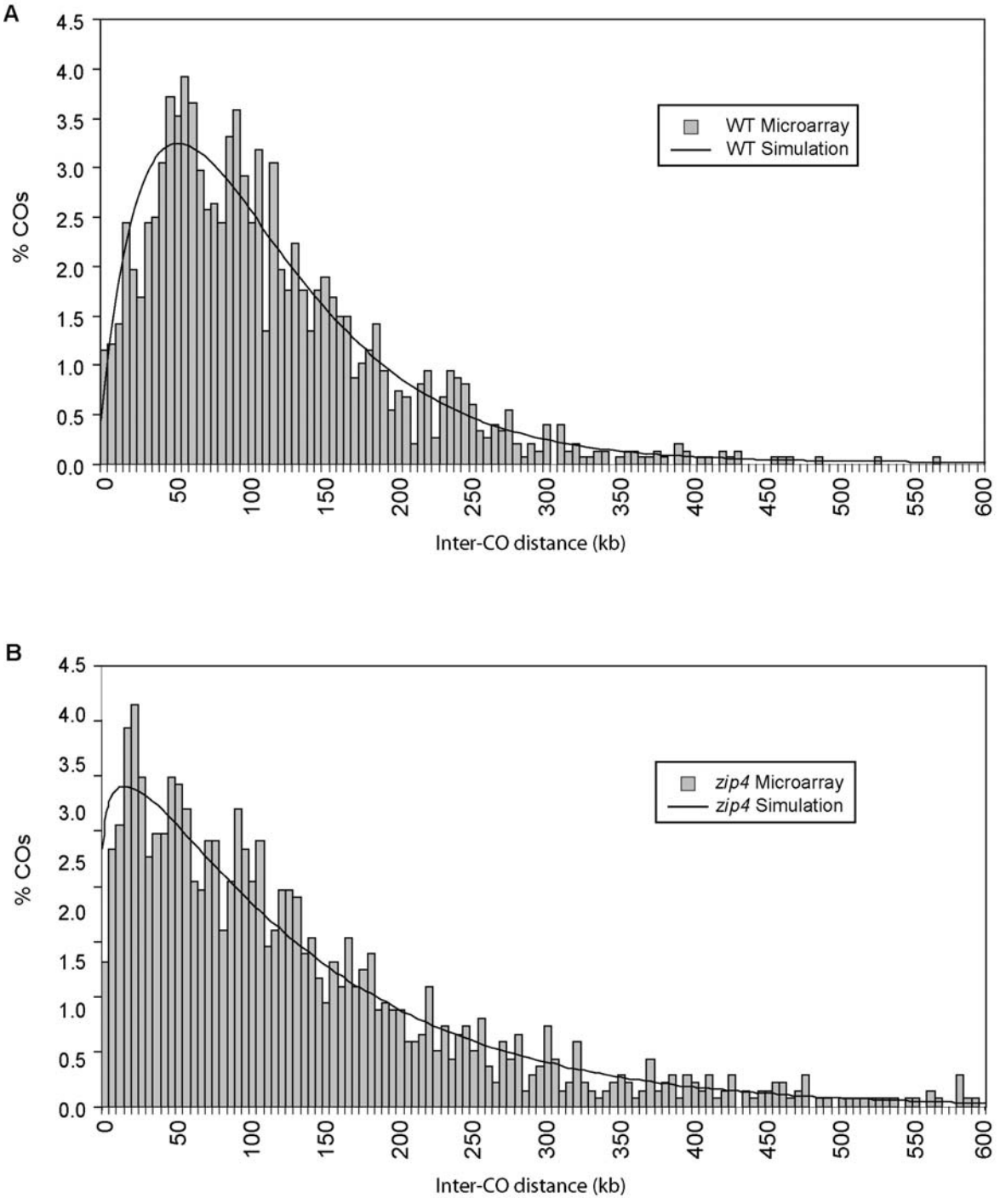

Figure S2. Distribution of Inter-crossover Distances for Wild type and *zip4*

Comparison of microarray and best-fit gamma distribution for inter-CO distances for wild type with normal interference (A) and *zip4* with reduced interference (B) as presented in Figure 4A and 4B, but with a smaller bin size of 5 kb. Microarray data is presented in gray bars; simulation data is plotted as a black line.

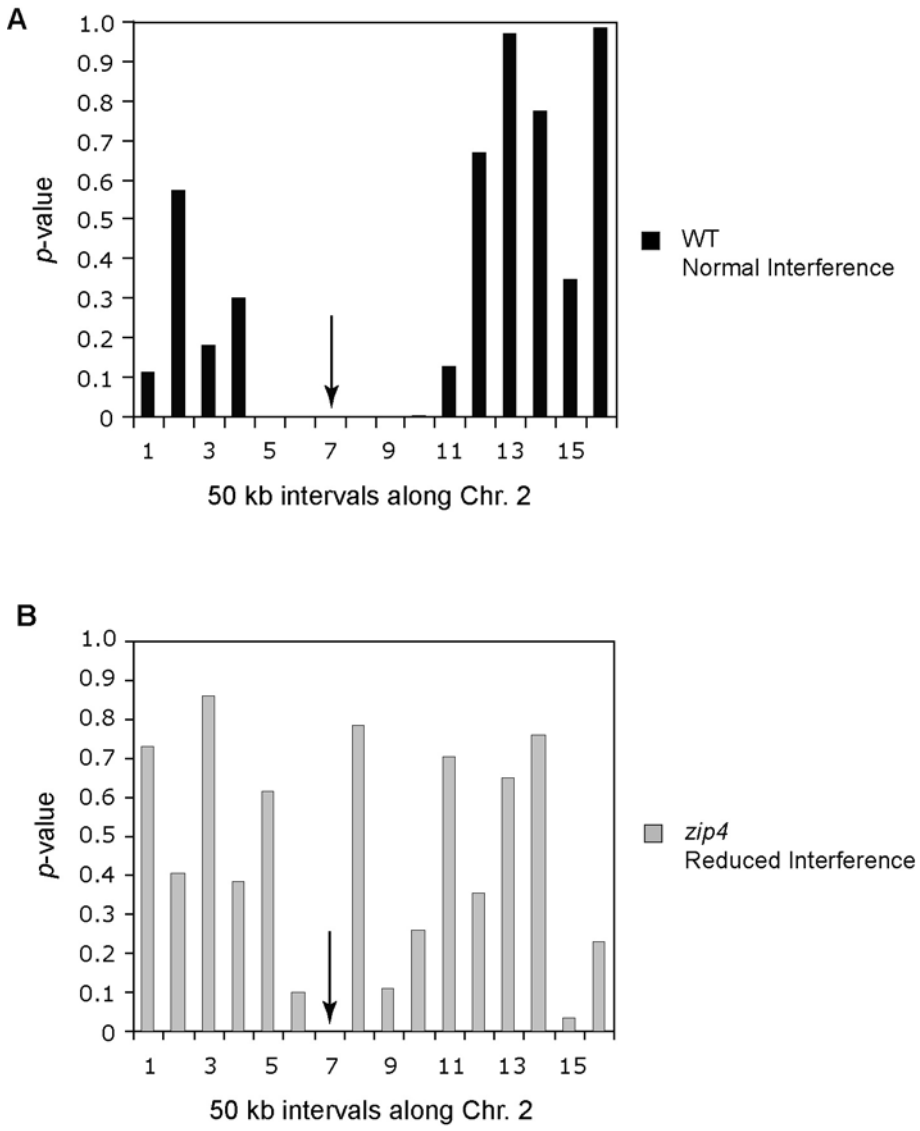

Figure S3. Analysis of Interference using the Malkova Method

CO interference is analyzed using the method presented in Malkova et al. (2004). CO distributions are simulated for 10000 meioses using the hazard functions with gamma values obtained from wild type and *zip4* microarray data. Each chromosome is divided into 50 kb intervals (~15 cM) and the number of COs per interval is determined and assigned to be a PD, TT or NPD. Each interval in turn becomes a reference interval and PD:TT:NPD ratios are calculated for each adjacent interval depending on whether a CO event (TT or NPD) or no COs (PD) had occurred. A chi-square test is then performed, in each of the adjacent intervals, to determine whether the difference between the PD:TT:NPD ratio of the two sample pools (with or without a CO in the reference interval) is significant. A high *p*-value indicates no difference between the two pools and suggests no effects of interference between the reference interval and the particular adjacent interval. A *p*-value lower than 0.05 indicates significant difference between the PD:TT:NDP ratio of the two sample pools, and suggests that interference extends from the reference interval into the particular adjacent interval. Shown here is a plot of *p*-values for a representative interval (7) on Chromosome 2, indicated by a vertical black arrow, and its adjacent intervals along the chromosome for WT (A) and *zip4* (B).

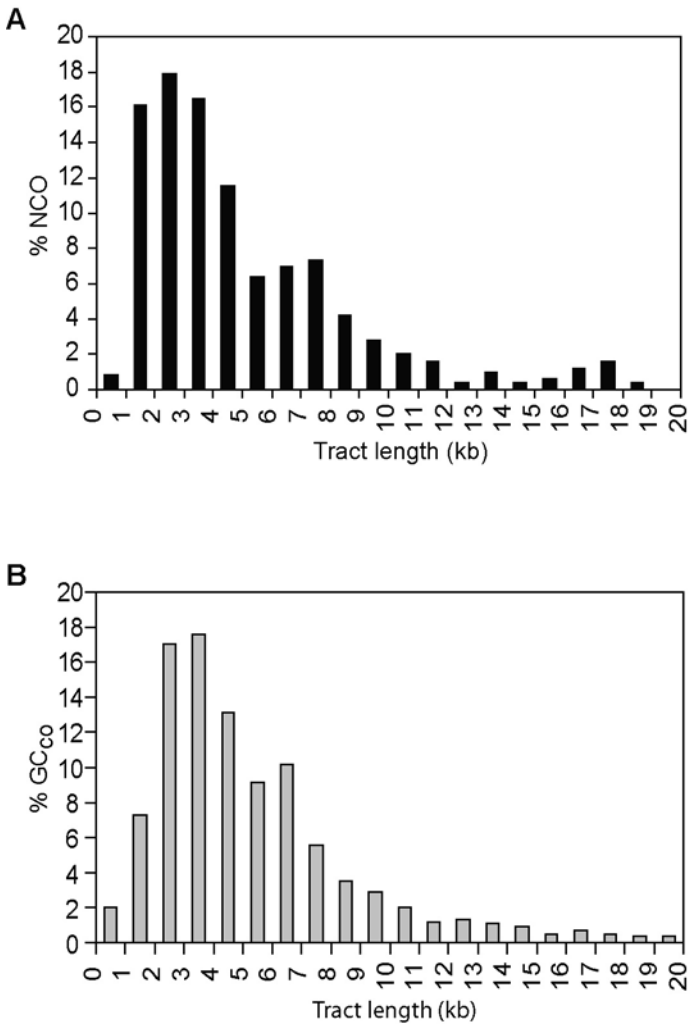

Figure S4. Distribution of GC Tract Lengths in Wild Type

Wild-type GC tract lengths from the microarray data plotted as a histogram with 1-kb bin size. (A) Tract length distribution of GCs unassociated with a CO (NCOs) (B) Tract length distribution of GCs associated with crossovers (GC_{CO}).

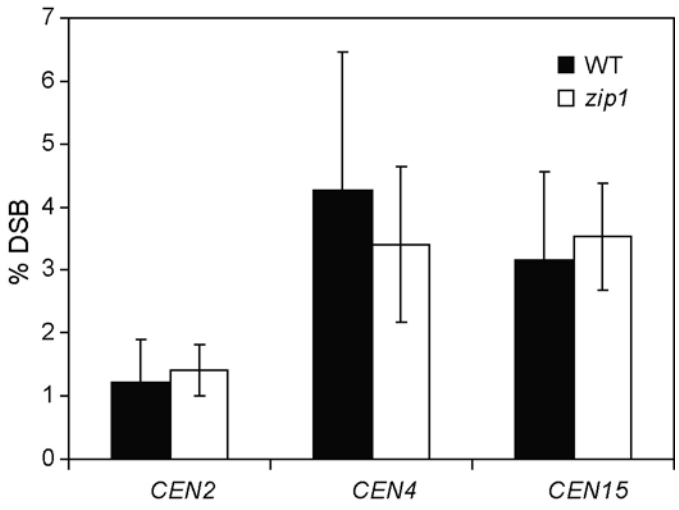

Figure S5. Quantification of Southern Blot Analysis of DSB Hotspots

Southern analysis of DSB hotspots near the centromeres in a *zip1 dmc1* background (as shown in Figure 6F) was quantified for wild type and the *zip1* mutant. For each hotspot, the average intensity obtained from the 5-hr and 8-hr time points was used. Hotspots YBL002W/3C, YDL004W/5C and YOL001W/2C are quantified for *CEN2*, *CEN4*, and *CEN5*, respectively. Error bars depict standard deviation.

Table S1. Spore Viability and Sporulation Frequency for Wild Type and Mutants

	Spore Viability (%)						Sporulation Frequency (%)	
	0 s.v.	1 s.v.	2 s.v.	3 s.v.	4 s.v.	total viability	4-spore asci	total sporulation
WT	0.5	3.5	13.8	37.3	44.9	80.6	14.6	36.9
zip1	25.0	32.0	25.6	13.0	4.3	34.9	0.4	5.4
zip2	20.6	18.8	28.5	16.7	15.5	46.9	16.1	38.7
zip3	30.5	23.7	29.8	11.5	4.6	34.0	8.8	25.8
zip4	27.5	2.5	40.0	17.5	12.5	46.3	7.5	17.9
msh4	30.1	14.8	23.8	16.2	15.2	42.9	5.6	12.6
spo16	14.6	10.0	24.0	33.3	18.1	57.6	17.9	33.7
sgs1	25.4	24.9	28.5	17.1	4.2	37.4	5.2	8.4
ndj1	32.8	22.7	22.7	16.4	5.5	34.8	4.1	16.2

Numbers of 0, 1, 2, 3, or 4 spore viable tetrads (s. v.) from tetrad dissection were counted, and the frequency of each type is reported. Sporulation frequencies were measured in the S96/YJM789 diploid strain background at least 72 hours after transferring onto plates containing sporulation medium. % total sporulation reports the frequency of cells that have completed MI or MI and MII, as determined by DAPI (4'-6'-Diamindino-2-phenylindole) staining of nuclei. All mutants show lower overall spore viability than wild type.

Table S2. Parity of Marker Segregation and NCO Frequencies

	S96:YJM789 Marker Segregation Frequency (%)					NCO Frequency (%)	
	0:4	1:3	2:2	3:1	4:0	S96	YJM789
WT	0.0	1.1	98.0	0.9	0.0	37.2	62.8
zip1	0.1	3.5	92.7	3.6	0.1	47.7	52.3
zip2	0.5	2.3	95.7	1.4	0.0	45.6	54.4
zip3	0.1	1.7	96.5	1.7	0.1	51.0	49.0
zip4	0.1	1.8	96.6	1.6	0.0	43.0	57.0
msh4	0.4	0.8	98.0	0.7	0.0	46.6	53.4
spo16	0.7	1.8	96.0	1.4	0.0	42.1	57.9
ndj1	0.0	1.5	96.6	1.6	0.3	38.4	61.6
sgs1	0.7	2.2	94.8	1.8	0.5	35.8	64.3

The parental origin (S96 or YJM789) of each marker was determined from each of the four spores making up one tetrad. The frequency of markers exhibiting one of the five marker segregation patterns (S96:YJM789: 0:4, 1:3, 2:2, 3:1, 4:0) was determined. As expected, most of the markers show a 2:2 segregation pattern. The frequency of NCOs, exhibiting a 3:1 or 1:3 marker segregation pattern and no associated CO, is also shown. In wild type, a greater proportion of NCOs show a 1:3 pattern of marker segregation, indicating that S96 sequences are converted to YJM789 sequences more often than YJM789 sequences are converted to S96. This proportion is maintained in *ndj1* and *sgs1*. However, less of a bias towards conversion to YJM789 is seen for the rest of the mutants.

Table S3. Comparison of NCO vs. GC_{co} Median Tract Lengths in Wild Type and Mutants

	WT	*zip1*	*zip2*	*zip3*	*zip4*	*msh4*	*spo16*	*sgs1*
Median NCO (kb)	3.9	5.1	4.7	4.6	5.0	4.2	5.3	4.1
Median GCco (kb)	4.4	6.3	5.8	5.8	5.7	5.8	6.4	4.8
χ^2	7.0	19.8	18.0	9.4	13.8	13.9	10.3	4.4
p-value	0.008	0.000	0.000	0.002	0.000	0.000	0.001	0.036
Conclusion	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

The median test (Mood, 1950) was used to test the hypothesis H_0 that the median tract lengths of NCOs vs. GC_{co} within a strain are the same. A grand median is determined from all the samples from both populations and tract lengths above and below it is tallied for both NCOs and GC_{co}s and compared using a 2x2 contingency table by chi-square analysis. *ndj1* was not tested because of insufficient sample size. Because the distribution of tract lengths is skewed, the median was chosen over the mean as a better measure of the central tendency. The significance level is set to 0.05. For all strains examined, the hypothesis that the GC tract lengths are the same for NCOs and GC_{co}s was rejected. Conversion tract lengths are larger in GCs associated with a CO than in NCOs.

Table S4. Genetic Measurements of CO Frequency and Interference

Strains	Background	Interval	P	T	NPD	NPD_{exp}	NPD ratio	Interference (1-NPD ratio)	cM
WT	BR1919 2n	*HIS4-LEU2*	615	378	7	25	0.28	0.72	21
WT	BR1919 2n	*LEU2-MAT*	519	475	15	44	0.34	0.66	28
zip2	BR1919 2n	*HIS4-LEU2*	2962	552	10	12	0.82	0.18	9
zip2	BR1919 2n	*LEU2-MAT*	2770	820	18	28	0.64	0.36	13
zip3	BR1919 2n	*HIS4-LEU2*	864	264	2	9	0.22*	0.78	12
zip3	BR1919 2n	*LEU2-MAT*	705	442	31	29	1.08	-0.08	27
spo16	BR1919 2n	*HIS4-LEU2*	1516	294	9	7	1.33	-0.33	10
spo16	BR1919 2n	*LEU2-MAT*	1431	423	5	14	0.35	0.65	12

P = parental ditype, T = tetratype, NPD = nonparental ditype and cM = centiMorgan. NPD_{exp} = number of NPDs expected in the absence of interference. The NPD ratio is the number of NPDs observed relative to the number of NPDs expected. Interference is given as 1-NPD ratio.

Table S5. Comparison of NCO Frequencies among Strains

Strain A vs. Strain B	NCO Mean of Strain A	NCO Mean of Strain B	q	$q'_{0.05,\infty,8}$	Conclusion (H_0: Mean A = Mean B)
zip1 vs. *msh4*	71	17	25.306	4.286	reject H_0
zip1 vs. WT	71	19	28.549	4.286	reject H_0
zip1 vs. *ndj1*	71	22	20.169	4.286	reject H_0
zip1 vs. *zip2*	71	30	22.620	4.286	reject H_0
zip1 vs. *spo16*	71	34	16.617	4.286	reject H_0
zip1 vs. *zip4*	71	35	20.877	4.286	reject H_0
zip1 vs. *zip3*	71	37	5.661	4.286	reject H_0
zip1. vs. *sgs1*	71	36	15.923	4.286	reject H_0
sgs1 vs. *msh4*	36	17	9.891	4.286	reject H_0
sgs1 vs. WT	36	19	10.712	4.286	reject H_0
sgs1 vs. *ndj1*	36	22	6.220	4.286	reject H_0
sgs1 vs. *zip2*	36	30	4.424	4.286	reject H_0
sgs1 vs. *spo16*	36	34	1.975	4.286	accept H_0
sgs1 vs. *zip4*	36	35	1.929	4.286	accept H_0
sgs1 vs. *zip3*	36	37	0.333	4.286	accept H_0
zip3 vs. *msh4*	37	17	8.591	4.286	reject H_0
zip3 vs. WT	37	19	8.958	4.286	reject H_0
zip3 vs. *ndj1*	37	22	5.375	4.286	reject H_0
zip3 vs. *zip2*	37	30	3.378	4.286	accept H_0
zip3 vs. *spo16*	37	34	1.384	4.286	accept H_0
zip3 vs. *zip4*	37	35	1.129	4.286	accept H_0
zip4 vs. *msh4*	35	17	10.229	4.286	reject H_0
zip4 vs. WT	35	19	12.268	4.286	reject H_0
zip4 vs. *ndj1*	35	22	5.634	4.286	reject H_0
zip4 vs. *zip2*	35	30	3.540	4.286	accept H_0
zip4 vs. *spo16*	35	34	0.632	4.286	accept H_0
spo16 vs. *msh4*	34	17	7.101	4.286	reject H_0
spo16 vs. WT	34	19	7.239	4.286	reject H_0
spo16 vs. *ndj1*	34	22	4.038	4.286	accept H_0
spo16 vs. *zip2*	34	30	1.714	4.286	accept H_0
zip2 vs. *msh4*	30	17	7.301	4.286	reject H_0
zip2 vs. WT	30	19	8.153	4.286	reject H_0
zip2 vs. *ndj1*	30	22	3.326	4.286	accept H_0
ndj1 vs. *msh4*	22	17	2.502	4.286	accept H_0
ndj1 vs. WT	22	19	1.943	4.286	accept H_0
WT vs. *msh4*	19	17	1.078	4.286	accept H_0

A Tukey multiple comparison test with unequal sample sizes was used to test the hypothesis that the average number of NCOs are the same for all the strains with a significance level of 0.05. This test performs pair-wise testing of strain A vs. strain B using means ranked in order of magnitude. A q-statistic is calculated and compared to the critical value, q'. The hypothesis that the mean NCO frequencies are the same between strain A vs. strain B is rejected if $q > q'$.

Table S6. Comparison of GC_{CO} and NCO Tract Length Medians between Strains

Strain A vs. Strain B	n	Median GCco Tract Length of Strain A (kb)	Median GCco Tract Length of Strain B (kb)	q	$q'_{0.05,5,\infty}$	Conclusion (H_0: Median A = Median B)
zip1 vs. *zip2*	434	6.4	5.7	1.63	3.86	accept H_0
zip1 vs. *zip4*	434	6.4	5.6	2.21	3.86	accept H_0
zip1 vs. *sgs1*	434	6.4	4.8	6.43	3.86	reject H_0
zip1 vs. WT	434	6.4	4.4	6.91	3.86	reject H_0
zip2 vs. *zip4*	434	5.7	5.6	0.58	3.86	accept H_0
zip2 vs. *sgs1*	434	5.7	4.8	4.80	3.86	reject H_0
zip2 vs. WT	434	5.7	4.4	5.28	3.86	reject H_0
zip4 vs. *sgs1*	434	5.6	4.8	4.22	3.86	reject H_0
zip4 vs. WT	434	5.6	4.4	4.70	3.86	reject H_0
sgs1 vs. WT	434	4.8	4.4	0.48	3.86	accept H_0
Strain A vs. Strain B	n	Median NCO Tract Length of Strain A (kb)	Median NCO Tract Length of Strain B (kb)	q	$q'_{0.05,5,\infty}$	Conclusion (H_0: Median A = Median B)
zip1 vs. *zip2*	400	5.1	4.7	2.30	3.86	accept H_0
zip1 vs. *zip4*	400	5.1	4.4	3.17	3.86	accept H_0
zip1 vs. *sgs1*	400	5.1	4.0	4.99	3.86	reject H_0
zip1 vs. WT	400	5.1	4.0	5.37	3.86	reject H_0
zip2 vs. *zip4*	400	4.7	4.4	0.96	3.86	accept H_0
zip2 vs. *sgs1*	400	4.7	4.0	2.78	3.86	accept H_0
zip2 vs. WT	400	4.7	4.0	3.17	3.86	accept H_0
zip4 vs. *sgs1*	400	4.4	4.0	1.82	3.86	accept H_0
zip4 vs. WT	400	4.4	4.0	2.21	3.86	accept H_0
sgs1 vs. WT	400	4.0	4.0	0.38	3.86	accept H_0
zip2 vs. *zip4*	503	4.7	4.7	0.45	3.31	accept H_0
zip2 vs. WT	503	4.7	3.9	4.64	3.31	reject H_0
zip4 vs. WT	503	4.7	3.9	5.97	3.31	reject H_0

A Tukey-type multiple comparison test of the median with equal samples was used to test the hypothesis that median GC_{CO} tract lengths are the same between the mutants to a significance level of 0.05. The same test was applied to the median NCO tract lengths. Since the test requires equal sample sizes, strains with larger sample sizes than the strain with the lowest sample size had their sample size equalized by selecting a random subpopulation of tract lengths from the larger pool. Comparisons were only made with strains with large enough sample sizes. For the median NCO tract lengths, an additional comparison was made specifically for those strains (wild type (WT), *zip2* and *zip4*) with larger samples sizes so that medians could be more accurately determined and compared. If $q > q'$, the hypothesis that the medians are same is rejected.

Table S7. Nonexchange Chromosomes

Chr #	Chr size (kb)	WT	*zip1*	*zip2*	*zip3*	*zip4*	*msh4*	*spo16*	*ndj1*	*sgs1*	Total
1	230	0	3	12	1	8	2	1	1	1	31
6	270	0	0	5	3	4	1	0	1	0	16
3	317	0	0	5	0	2	1	1	1	0	10
9	440	0	0	3	1	1	2	0	0	0	8
8	563	0	0	2	1	2	0	0	0	0	5
5	577	0	0	2	0	2	1	0	0	0	5
11	666	0	0	0	1	2	1	0	0	0	5
10	746	0	0	0	0	1	0	1	0	0	2
14	784	0	0	1	1	3	1	0	1	0	7
2	813	0	0	0	0	3	0	0	0	0	3
13	924	0	0	0	0	0	0	1	0	0	1
16	948	0	0	1	0	1	0	0	0	0	2
12	1078	0	0	5	0	0	0	0	0	0	5
7	1091	0	0	0	0	1	0	0	0	0	1
15	1091	0	0	0	0	0	1	0	0	0	1
4	1532	0	0	0	0	0	0	0	0	0	0
# Tetrads		26	9	26	8	34	11	8	7	7	
% E_0		0.0	2.1	8.7	6.3	5.5	5.7	3.9	6.3	2.7	

Number of E_0s (chromosomes (chr) without a CO) found for each of the 16 chromosomes in wild type and all mutants. Chromosomes are arranged in order of increasing size. Total number of E_0s over all strains were tallied. A trend is seen that smaller chromosomes show a higher incidence of E_0s. % E_0 is the incidence of E_0s divided by the total number of chromosomes for all the tetrads in one strain background.

Table S8. Analysis of the Effects of the *zip4* Outlier Tetrad

		CO Count		Interference	NCO Count		Chromatid Interference	
	# Tetrad	Mean	SD	NPD ratio	Mean	SD	2-s.d : 3-s.d. : 4-s.d. Ratio	*p*-value
zip4 (all data)	34	56.7	17	0.96	36.7	16.3	1.0 : 2.0 : 1.1	0.45
zip4 (no outlier)	33	54.5	12	0.90	34.7	11.8	1:0 : 1.9 : 1.1	0.43

One *zip4* tetrad was found to have a much higher number of crossovers (126 COs) than remaining 33 *zip4* tetrads. To determine the effects of this outlier, we analyzed the *zip4* data set with and without the outlier. Presented here is the comparison of CO count, interference, NCO count, and chromatid interference for *zip4* data with and without the outlier tetrad.

Table S9. Yeast Strains

Strain	Genotype
S96	*MAT***a** *ho lys5*
YJM789	*MATα ho::hisG lys2 cyh*
JCF1850	S96 but *zip1::kanMX6*
JCF1852	YJM789 but *zip1::kanMX6*
JCF1104	S96 but *zip2::kanMX6*
JCF1106	YJM789 but *zip2::kanMX6*
SYC1104	S96 but *zip3::kanMX6*
SYC1105	YJM789 but *zip3::kanMX6*
TY461	S96 but *zip4::kanMX6*
TY462	YJM789 but *zip4::kanMX6*
SYC1110	S96 but *msh4::kanMX6*
SYC1111	YJM789 but *msh4::kanMX6*
JCF2035	S96 but *spo16::kanMX6*
JCF1210	YJM789 but *spo16::kanMX6*
SYC1120	S96 but *sgs1::kanMX6*
SYC1121	YJM789 but *sgs1::kanMX6*
SYC1112	S96 but *ndj1::kanMX6*
SYC1113	YJM789 but *ndj1::kanMX6*
BR4633	*leu2::CUP1,arg4-8 iTHR1 iura3-1 MATα iADE2* *leu2::CUP1,arg4-8 iura3-stu iNAT iLEU2 MAT***a** *trp1-289 ade2-1 ura3-1* *trp1-289 ade2-1 ura3-1*
BR4790	BR4633 but *zip1::kanMX6*
BR4829	BR4633 but *zip2::kanMX6*
S2937	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1*
JCF406	BR1919-8B *MAT***a** *ura3-1 trp1-289 thr1-4 ade2-1*
S1561	BR1919-8B *MATα leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1*
JS03	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1 zip2::URA3* *MATα ura3-1 trp1-289 thr1-4 ade2-1 zip2::URA3*
BR3643	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1 zip3::URA3* *MATα ura3-1 trp1-289 thr1-4 ade2-1 zip3::URA3*
JS36	BR1919-8B *MAT***a** *leu2-3,112 his4-260 ura3-1 trp1-289 thr1-4 ade2-1 spo16::kanMX6* *MATα ura3-1 trp1-289 thr1-4 ade2-1 spo16::kanMX6*
NKY1455	SK1 *MAT***a** *leu2::hisG his4X-LEU2-URA3 ura3 ho::LYS2 lys2 arg4-nsp dmc1::ARG4* *MATα leu2::hisG his4B-LEU2 ura3 ho::LYS2 lys2 arg4-bgl dmc1::ARG4*
YAH2650	Same as NKY1455, but *zip1::LYS2/zip1::LYS2*

Supplemental Procedures

Simulations

We assume that the inter-crossover distance distribution can be described by a gamma distribution, characterized by a shape parameter γ and a scale parameter β, according to the following probability density function.

$$f(x)=\frac{\left(\frac{x}{\beta}\right)^{\gamma-1}e^{-\frac{x}{\beta}}}{\beta\Gamma(\gamma)}$$

The corresponding cumulate distribution function is then

$$F(x)=\int_0^x \frac{\left(\frac{t}{\beta}\right)^{\gamma-1}e^{-\frac{t}{\beta}}}{\beta\Gamma(\gamma)}dt$$

letting u = t/β, and dt = β du, we obtain

$$F(x)=\frac{1}{\beta\Gamma(\gamma)}\int_0^{x/\beta} u^{\gamma-1}e^{-u}\beta du$$

$$=\frac{1}{\Gamma(\gamma)}\int_0^{x/\beta} u^{\gamma-1}e^{-u}du$$

$$=\frac{\Gamma_{(x/\beta)}(\gamma)}{\Gamma(\gamma)}$$

where the numerator is a partial gamma function evaluated for x/β.

The hazard function is defined as the function giving the probability that, given an event (in this case, a CO) at position 0, a second event will occur at position x. The hazard function is equal to the probability density function divided by 1 minus the cumulative distribution function. Therefore,

$$h(x) = \frac{f(x)}{1 - F(x)}$$

$$= \frac{\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{x}{\beta}}}{\beta\Gamma(\gamma)} \bullet \frac{1}{1 - \frac{\Gamma_{\left(x/\beta\right)}(\gamma)}{\Gamma(\gamma)}}$$

$$= \frac{\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{x}{\beta}}}{\beta\left(\Gamma(\gamma) - \Gamma_{\left(x/\beta\right)}(\gamma)\right)}$$

Evaluation of hazard function h(x) therefore requires computation of both a complete gamma function as well as a partial gamma function. To use this in the simulation, the current array describing the probability of placing the next CO at any given position is to be multiplied by the hazard function centered at the position of the most recently placed CO. Initially, the array is set so the probability of getting a CO is uniform across the entire genome. The simulation does not take into account CO hotspots. Hotspots were not incorporated due to lack of high resolution data in which the relative strengths of hotspots are known on a genome-wide basis. As COs are added sequentially, the probabilities in the array are modified by the gamma-based interference function as described below. For a gamma distribution with gamma greater than or equal to one, the hazard function is small (near zero) and then increases smoothly until it approaches one. Multiplication by this function will essentially remove a portion of the probability density function around the recently placed CO, thus decreasing the likelihood that subsequent COs will occur near that position.

The parameters of the gamma distribution can be obtained directly from the raw data (list of inter-crossover distances) using the moment estimators (Evans M et al., 2000). An improved moment estimator with correction for small sample size (Hwang TY and Huang PH, 2002) was tested and found to yield no difference for the data used here. We therefore used the standard moment estimators:

$$\hat{\gamma} = \left(\frac{\bar{x}}{s}\right)^2$$

$$\hat{\beta} = \frac{s^2}{\bar{x}}$$

Both the average number of COs obtained from the microarray data and its variance are used to determine the total number of COs simulated.

To calculate simulated NPD ratios, the genome was divided into equal intervals and for the frequencies of PDs, TTs and NPDs were tallied from the simulated crossover distributions. 10,000 meioses were simulated for each NPD analysis. Since the value of the NPD ratio varies as a function of the interval size, the interval size was chosen as the average of the cM size used in the published genetic determinations of NPD ratios in order to best compare genetic vs. microarray NPD ratios. To obtain average simulated NPD ratios, the NPD ratio for each interval was averaged together for all intervals.

Supplemental References

Evans M, Hastings N, and B., P. (2000). Statistical Distributions, 3rd edn (New York, Wiley).

Hunter, N., Chambers, S.R., Louis, E.J., and Borts, R.H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J *15*, 1726-1733.

Hwang TY, and Huang PH (2002). On new moment estimation of parameters of the gamma distribution using its characterization. Ann Inst Statist Math *54*, 840-847.

Malkova, A., Swanson, J., German, M., McCusker, J.H., Housworth, E.A., Stahl, F.W., and Haber, J.E. (2004). Gene conversion and crossing over along the 405-kb left arm of Saccharomyces cerevisiae chromosome VII. Genetics *168*, 49-63.

Mood, A.M. (1950). Introduction to the Theory of Statistics (New York, McGraw-Hill).