Supplemental Figures

Figure S1: Defining intrachromosomal breakpoints. **A:** For *S. cerevisiae* intergenic regions flanked by *K. waltii* homologs located on the same *K. waltii* chromosome, the number of genes between the *K. waltii* homologs in their native *K. waltii* gene order was counted. **B:** The distribution of the number of genes between the *K. waltii* homologs flanking an *S. cerevisiae* intergenic region. The inset shows the histogram for *S. cerevisiae* intergenic regions flanked by *K. waltii* homologs located 10 or more genes apart. Intergenic regions with *K. waltii* homologs 20 genes (red line) or more apart were called as intrachromosomal breakpoints. The data for Wolfe *S. cerevisiae – K. waltii* intergenic regions are shown. Similar results were found for Kellis *S. cerevisiae – K. waltii* as well as for Wolfe *S. cerevisiae* – Ancestor and Wolfe *K. waltii* – Ancestor.

Figure S2: Distances between adjacent breakpoints. For the enrichment and regression analyses, breakpoints within one bin are treated as a single breakpoint and therefore clustered breakpoints become a single event. To decide on the optimal bin size to ensure that clustered breakpoints end up in the same bin, we plotted the distribution of distances between adjacent breakpoints. The histogram for the Wolfe *S. cerevisiae – K. waltii* genome comparison is shown here. Similar histograms were generated for the other genome comparisons. From these histograms, we noted a predominance of breakpoints within 5 kb of each other and decided to use a bin size of 5 kb (red line) for the enrichment and regression analyses. Over 80% of the breakpoints found by any genome comparison that lie within 5 kb of one another end up in the same bin when 5 kb bins are used.

Figure S3: Size distribution of all *S. cerevisiae* intergenic regions and those bearing a breakpoint for the Wolfe *S. cerevisiae – K. waltii* genome comparison. The inset shows the histogram for intergenic sizes of 12 kb or greater. Similar histograms were generated for Kellis *S. cerevisiae – K. waltii* and Wolfe *S. cerevisiae* – Ancestor comparisons. From both the Wolfe and Kellis *S. cerevisiae – K. waltii* plots, we noticed that above 17 kb (the red line on the inset), all or all but one intergenic regions bear breakpoints. Thus, to exclude the largest intergenic regions and hence also the largest breakpoints, we used a cutoff of 17 kb for the minimal endpoint distance analysis. A similar histogram was created for the Wolfe *K. waltii* - Ancestor genome comparison and a 14 kb cutoff was deemed appropriate.

Supplemental Tables

Table S1: Numbers of genes for *S. cerevisiae*, *K. waltii*, and the Ancestor and the number of homologs between compared species for the Kellis dataset (Kellis et al. 2004) and the Wolfe dataset (Byrne et al. 2005) (http://wolfe.gen.tcd.ie/ygob/). For the list of genes used see Resources and Datasets.

	Kellis dataset	Wolfe dataset
S. cerevisiae genes	6328	5615
K. waltii genes	5060	5060
Ancestor genes	NA	4699
S. cerevisiae genes with a *K. waltii* homolog	5146	4906
S. cerevisiae genes with an Ancestor homolog	NA	5105
K. waltii genes with an Ancestor homolog	NA	4424

Table S2: Genomic features used to search for correlations with breakpoints. The features were derived from: [1]SGD (http://www.yeastgenome.org/); [2]Borde et al. (2004); [3]Nieduszynski et al. (2007); [4]this work; [5]McCune et al. (2008); [6]Kellis et al. (2004). Telomeric repeats are those features annotated in SGD as "repeat_family" while LTRs are features denoted as "repeat_region". Tys comprise features listed as either "transposable_element" or "transposable_element_gene" in SGD. Therefore, full Tys may be annotated as a complete feature or as multiple separate genes. Ribosomal protein genes are named rpl*, rps*, mrlp*, or mrsp* in SGD and include both the cytoplasmic and mitochondrial small and large ribosomal subunit genes. The list of all genomic features used is provided in Resources and Datasets.

Genomic Feature	Total Number
S. cerevisiae centromeres[1]	16
S. cerevisiae telomeres[1]	32
S. cerevisiae telomeric repeats[1]	100
S. cerevisiae snoRNAs[1]	77
S. cerevisiae snRNAs[1]	6
S. cerevisiae ribosomal genes[1]	183
S. cerevisiae Spo11 hotspots[2]	574
S. cerevisiae tRNAs[1]	275
S. cerevisiae Tys[1]	139
S. cerevisiae LTRs[1]	392
S. cerevisiae OriDB ARSs[3]	732
Confirmed	326
Likely	244
Dubious	165
S. cerevisiae high confidence ARSs (hcARSs)[4]	411
S. cerevisiae McCune origins[5]	200
NonCDR (early firing)	77
CDR (late firing)	123
Rad53 unregulated (early firing)	101
Rad53 regulated (late firing)	99
K. waltii centromeres[6]	8
K. waltii tRNAs[4]	217

Table S3: Number and types of intergenic regions among *S. cerevisiae*, *K. waltii*, and the Ancestor. The type of intergenic region is determined according to if and how the flanking genes overlap irrespective of reading frame or coding strand. Type A: the flanking genes overlap such that the first gene ends before the second. Type B: the flanking genes overlap such that the first gene ends after the second. Type C: the flanking genes do not overlap. Only Type C intergenic regions were interrogated as potential breakpoints in this study. Counts for the three types of intergenic regions are given for intergenic regions found for *S. cerevisiae* – *K. waltii*, *S. cerevisiae* – Ancestor, and *K. waltii* – Ancestor for the Kellis and Wolfe datasets.

Intergenic Type	Kellis *S. cerevisiae* - *K. waltii*	Wolfe *S. cerevisiae* - *K. waltii*	Wolfe *S. cerevisiae* - Ancestor	Wolfe *K. waltii* - Ancestor
A 	65	9	10	46
B 	2	0	0	16
C 	5063	4881	5079	4354

Table S4: Literature references from which 442 breakpoints were curated. Plain-text references (seven papers, 147 breakpoints) report evolutionary breakpoints among *sensu stricto* yeast. References in bold (22 papers, 295 breakpoints) represent those in which the authors generated the breakpoints experimentally. Note that in the HO break assay (VanHulle et al. 2007), though the site of the HO break is engineered and known, the site to which the HO break repairs represents a biologically determined break and repair event.

Admire et al. (2006)	Dunn et al. (2005)	Kellis et al. (2003)	Rachidi et al. (1999)
Aladjem et al. (1988)	Fischer et al. (2000)	**Koszul et al. (2004)**	**Schacherer et al. (2004)**
Argueso et al. (2008)	**Gresham et al. (2008)**	**Lemoine et al. (2005)**	**Schacherer et al. (2005)**
Bond et al. (2004)	Hall et al. (2005)	**Lemoine et al. (2008)**	**Schacherer et al. (2007)**
Brown et al. (1998)	**Huang and Koshland (2003)**	**Maxwell and Curcio (2007)**	**Tourrette et al. (2007)**
Chen and Kolodner (1999)	**Hwang et al. (2008)**	**Payen et al. (2008)**	**Umezu et al. (2002)**
Degtyareva et al. (2008)	**James et al. (2008)**	Perez-Ortinz et al. (2002)	**VanHulle et al. (2007)**
Dunham et al. (2002)			

Table S5: Number of breakpoints identified for each set of compared genomes. The total numbers of intergenic regions for the compared genomes are given. The subset of intergenic regions bearing each type of breakpoint is subsequently shown followed by the total number of breaks for each genome comparison. Numbers in parentheses represent values used for the minimal endpoint distance analysis. Lists of breakpoints for each dataset are provided in Resources and Datasets.

	Kellis *S. cerevisiae* – *K. waltii*	Wolfe *S. cerevisiae* – *K. waltii*	Wolfe *S. cerevisiae* - Ancestor	Wolfe *K. waltii* - Ancestor
Intergenic Regions	5063 (5059)	4881 (4875)	5079 (5077)	4354 (4345)
Interchr Breaks	827 (824)	496 (491)	139 (138)	153 (145)
Intrachr Breaks	162 (161)	114 (114)	57 (57)	40 (39)
Tandem Duplications	40 (40)	0 (0)	0 (0)	0 (0)
Gene Order Changes	123 (123)	108 (108)	85 (85)	76 (76)
Total Breaks	1152 (1148)	718 (713)	281 (280)	269 (260)

Table S6: Enrichment analysis of *S. cerevisiae – K. waltii* breakpoints and genomic features. Features and breakpoints were assigned a 5 kb bin by midpoint. *P*-values shown reflect the probability of seeing at least the observed co-localization of features and breakpoints under a hypergeometric distribution. Significant *P*-values ($P < 0.05$) are highlighted in yellow, and the entire box is highlighted in pink for those significant after a Bonferroni correction ($P < 0.0025$). Ty features used in this analysis are redundant: a single full length Ty could be represented as both one feature and as the individual genes comprising the Ty. As a result, a breakpoint falling inside a complete Ty may be placed in the same bin as the complete feature but will only be in the same bin as one of the Ty genes. This redundancy reduces the significance of the correlation between Tys and breakpoints. LTRs better represent the location of Ty elements than full Ty elements for this analysis.

Genomic Feature	Total number of bins with feature	Number of bins with feature and a Kellis breakpoint	Kellis *P*-value	Number of bins with feature and a Wolfe breakpoint	Wolfe *P*-value
Centromeres	16	6	0.5848	5	0.3925
Ribosomal protein genes	170	60	0.7320	45	0.4279
snoRNAs	60	23	0.4779	20	0.1103
snRNAs	5	1	0.9028	0	1.0000
Spo11 hotspots	409	139	0.9392	104	0.5628
Telomeres	32	6	0.9934	0	1.0000
Telomeric repeats	51	19	0.5517	2	1.000
tRNAs	254	125	$<10^{-4}$	105	$<10^{-8}$
Tys	65	20	0.8903	15	0.7282
LTRs	255	110	0.0236	87	0.0009
hcARSs	398	185	$<10^{-4}$	123	0.0055
OriDB ARSs	699	289	0.0045	182	0.4033
Confirmed	316	144	0.0007	93	0.0572
Likely	241	101	0.0657	57	0.7926
Dubious	163	51	0.9574	38	0.7851
McCune origins	200	90	0.0113	70	0.0013
nonCDR (early firing)	77	38	0.0182	28	0.0224
CDR (late firing)	123	52	0.1374	42	0.0192
Rad53 unregulated (early firing)	101	52	0.0020	41	0.0005
Rad53 regulated (late firing)	99	38	0.4423	29	0.2288

Table S7: Simulation analysis using minimal endpoint and midpoint measures for *S. cerevisiae* – *K. waltii* breakpoints and genomic features. Simulations were not attempted for features restricted to particular regions of the genome (centromeres, telomeres) or for rare features (snRNAs). [1]The number of breakpoints with a feature within 1 kb. [2]The mean number of breakpoints with a feature within 1 kb from 10,000 sets of simulated breakpoints. [3]*P*-values were determined by summing the number of simulations in which the number of breakpoints within 1 kb of a feature was equal to or greater than in the real data and dividing by the total number of simulations (10,000). Significant *P*-values ($P < 0.05$) are highlighted in yellow, and the entire box is highlighted in pink for those significant after a Bonferroni correction ($P < 0.0029$).

	Minimal Endpoint Distance Measure						Midpoint Distance Measure					
	Kellis *Scer* - *Kwal* break			Wolfe *Scer* - *Kwal* break			Kellis *Scer* - *Kwal* break			Wolfe *Scer* - *Kwal* break		
Feature	Observed ≤ 1kb[1]	Mean[2]	*P*-value[3]	Observed ≤ 1kb	Mean	*P*-value	Observed ≤ 1kb	Mean	*P*-value	Observed ≤ 1kb	Mean	*P*-value
Breakpoints	322	327.6	0.6260	111	125.8	0.8987	41	109.9	1.0000	12	42.7	1.0000
Ribosomal protein genes	69	78.3	0.8868	48	50.0	0.6367	51	62.8	0.9496	37	40.6	0.7500
snoRNAs	21	19.5	0.4021	18	12.4	0.0743	17	14.8	0.3055	13	9	0.1043
Spo11 hotspots	169	193.1	0.9573	118	124.6	0.7503	113	140.1	0.9908	83	90.9	0.8263
tRNAs	127	66.8	0.0001	107	41.9	0.0003	77	40.9	0.0001	65	25.6	0.0001
Tys	22	10.7	0.0003	17	6.5	$<10^{-4}$	16	7.7	0.0016	15	4.9	0.0001
LTRs	98	45.5	0.0003	76	27.5	$<10^{-4}$	59	25.7	0.0004	42	14.7	$<10^{-4}$
hcARSs	204	136.9	$<10^{-4}$	130	83.5	0.0002	124	83.8	0.0001	76	50.4	0.0002
OriDB ARSs	347	244.8	$<10^{-4}$	210	147.9	0.0002	180	135.9	0.0002	92	78.9	0.0532
Confirmed	142	89.6	0.0002	90	55.2	$<10^{-4}$	92	62.2	0.0001	52	36.1	0.0026
Likely	144	95.3	0.0005	81	55.3	0.0007	65	45.9	0.0016	26	25.5	0.4872
Dubious	78	71.0	0.2040	54	44.5	0.0794	23	27.9	0.8754	14	17.3	0.8344
McCune origins	97	72.6	0.0011	73	47.1	0.0003	59	42.8	0.0043	38	26.5	0.0109
nonCDR (early firing)	51	27.9	0.0003	36	18.1	0.0004	32	17.6	0.0005	20	10.7	0.0050
CDR (late firing)	46	44.6	0.4355	37	29.1	0.0729	27	25.3	0.3897	18	15.7	0.3045
Rad53 unregulated (early firing)	57	35.2	0.0002	44	22.6	0.0001	32	19.5	0.0021	21	12.2	0.0089
Rad53 regulated (late firing)	40	37.5	0.3735	29	24.6	0.2005	27	23.2	0.2192	17	14.3	0.2606

Table S8: Enrichment and simulation analysis using minimal endpoint and midpoint measures for Wolfe *S. cerevisiae* – Ancestor breakpoints and genomic features. The Ancestor is an inferred ancestor of *S. cerevisiae* and *K. waltii*. Its gene content was determined by parsimony using eleven yeast genomes (Gordon et al. 2009). The Ancestor genome used in this work is a previous version based off of eight yeast genomes. For details see the legends for **Tables S6** and **S7**.

	Enrichment Test			Minimal Endpoint Distance Measures Simulation Test			Midpoint Distance Measures Simulation Test		
Genomic Feature	Total number of bins with feature	Number of bins with feature and breakpoint	*P*-value	Observed ≤ 1kb	Mean	*P*-value	Observed ≤ 1kb	Mean	*P*-value
Breakpoints	NA	NA	NA	34	28.5	0.2105	8	10.6	0.7949
Centromeres	16	2	0.5121	NA	NA	NA	NA	NA	NA
Ribosomal protein genes	170	16	0.7202	16	21.6	0.9209	13	17.7	0.9036
snoRNAs	60	7	0.4430	6	4.8	0.3525	4	3.7	0.5090
snRNAs	5	0	1.0000	NA	NA	NA	NA	NA	NA
Spo11 hotspots	409	40	0.7215	43	48.6	0.8191	29	36	0.9057
Telomeres	32	0	1.0000	NA	NA	NA	NA	NA	NA
Telomeric repeats	51	0	1.0000	NA	NA	NA	NA	NA	NA
tRNAs	254	58	$<10^{-9}$	60	16	0.0001	38	10.1	0.0002
Tys	65	10	0.1363	12	2.4	0.0002	10	1.9	0.0003
LTRs	255	43	0.0006	43	10.2	$<10^{-4}$	23	5.7	$<10^{-4}$
hcARSs	398	57	0.0052	55	31.1	0.0002	35	19.6	0.0007
OriDB ARSs	699	77	0.3162	88	55.8	0.0001	51	31.4	0.0004
Confirmed	316	43	0.0355	37	20.8	0.0007	24	14.5	0.0086
Likely	241	26	0.4670	31	20.4	0.0164	18	9.8	0.0092
Dubious	163	11	0.9658	25	17.3	0.0498	9	7.3	0.3049
McCune origins	200	28	0.0616	28	17.8	0.0114	16	10.5	0.0603
nonCDR (early firing)	77	18	0.0007	20	6.9	0.0001	12	4.3	0.0008
CDR (late firing)	123	10	0.8483	8	10.8	0.8512	4	6.2	0.8731
Rad53 unregulated (early firing)	101	21	0.0013	10	8.5	0.0009	12	4.9	0.0036
Rad53 regulated (late firing)	99	7	0.9093	8	9.2	0.7037	4	5.6	0.8149

Table S9: Enrichment and simulation analysis using midpoint and minimal endpoint measures for Wolfe *K. waltii* – Ancestor breakpoints and genomic features. Only centromeres (Kellis et al. 2004) and tRNAs (this work) have been mapped in *K. waltii*. No data for origin locations in *K. waltii* are yet available. Significant *P*-values ($P < 0.05$) are highlighted in yellow, and the entire box is highlighted in pink for those significant after a Bonferroni correction ($P < 0.0250$).

	Enrichment Test			Minimal Endpoint Distance Measures Simulation Test			Midpoint Distance Measures Simulation Test		
Genomic Feature	Total number of bins with feature	Number of bins with feature and breakpoint	*P*-value	Observed ≤ 1kb	Mean	*P*-value	Observed ≤ 1kb	Mean	*P*-value
Breakpoints	NA	NA	NA	62	40.8	0.0012	8	15.7	0.9650
Centromeres	8	1	0.6289	NA	NA	NA	NA	NA	NA
tRNAs	201	35	0.0069	35	14.6	0.0005	27	12.2	0.0005

Table S10: Enrichment and simulation analysis using minimal endpoint measures of evolutionarily-derived breakpoints from the literature with genomic features. For details see the legends in **Tables S6** and **S7**.

	Enrichment Test			Minimal Endpoint Distance Measures Simulation Test		
Genomic Feature	Total number of bins with feature	Number of bins with feature and a breakpoint	*P*-value	Observed ≤ 1kb	Mean	*P*-value
Evolutionary literature breaks	NA	NA	NA	48	47.9	0.9555
Centromeres	0	16	1.0000	NA	NA	NA
Ribosomal protein genes	10	170	0.3315	13	10.4	0.2607
snoRNAs	0	60	1.0000	1	3.3	0.9456
snRNAs	0	5	1.0000	NA	NA	NA
Spo11 hotspots	17	409	0.8255	26	25.0	0.4490
Telomeres	2	32	0.4762	NA	NA	NA
Telomeric repeats	5	51	0.1050	NA	NA	NA
tRNAs	41	254	$<10^{-12}$	48	13.2	$<10^{-4}$
Tys	7	65	0.0392	8	4.7	0.1254
LTRs	40	255	$<10^{-11}$	43	13.8	0.0002
hcARSs	35	398	0.0002	36	25.0	0.0198
OriDB ARSs	53	699	0.0002	55	44.5	0.0518
Confirmed	28	316	0.0011	26	17.5	0.0411
Likely	20	241	0.0128	24	17.3	0.0921
Dubious	8	163	0.5660	13	13.2	0.5483
McCune origins	19	200	0.0036	21	12.8	0.0269
nonCDR (early firing)	10	77	0.0040	9	5.0	0.0871
CDR (late firing)	9	123	0.1517	12	7.8	0.1182
Rad53 unregulated (early firing)	14	101	0.0003	15	6.4	0.0059
Rad53 regulated (late firing)	5	99	0.5497	6	6.4	0.5926

Table S11: Enrichment and simulation analysis using minimal endpoint measures of experimentally derived breakpoints from the literature with genomic features. For details see the legends in **Tables S6** and **S7**.

	Enrichment Test			Minimal Endpoint Distance Measures Simulation Test		
Genomic Feature	Total number of bins with feature	Number of bins with feature and a breakpoint	*P*-value	Observed ≤ 1kb	Mean	*P*-value
Experimental literature breaks	NA	NA	NA	164	164	0.9079
Centromeres	16	2	0.3954	NA	NA	NA
Ribosomal protein genes	170	7	0.9924	10	26	0.9971
snoRNAs	60	7	0.2370	8	8.7	0.5647
snRNAs	5	1	0.3560	NA	NA	NA
Spo11 hotspots	409	29	0.8783	57	59.3	0.5851
Telomeres	32	2	0.7657	NA	NA	NA
Telomeric repeats	51	2	0.9378	NA	NA	NA
tRNAs	254	49	$<10^{-8}$	126	33.9	$<10^{-4}$
Tys	65	31	$<10^{-16}$	59	11.2	0.0011
LTRs	255	72	$<10^{-23}$	130	33.8	$<10^{-4}$
hcARSs	398	43	0.0408	88	62.1	0.0060
OriDB ARSs	699	68	0.0825	145	107	0.0007
Confirmed	316	37	0.0187	70	44.2	0.0036
Likely	241	21	0.4688	52	41.6	0.1434
Dubious	163	14	0.5119	47	33.1	0.0576
McCune origins	200	19	0.3205	47	32.4	0.0395
nonCDR (early firing)	77	9	0.1951	25	12.4	0.0220
CDR (late firing)	123	10	0.5971	22	20	0.3756
Rad53 unregulated (early firing)	101	13	0.0774	40	16.4	0.0001
Rad53 regulated (late firing)	99	6	0.8546	7	16	0.9709

Table S12: Enrichment and simulation tests for co-localization of experimentally generated breakpoints with a subset of *S. cerevisiae – K. waltii* evolutionary breakpoints. The *S. cerevisiae – K. waltii* breakpoints used here consist of those having a given feature within 1 kb by minimal endpoint distance measure. Significant *P*-values ($P < 0.05$) are highlighted in yellow, and the entire box is highlighted in pink for those significant after a Bonferroni correction ($P < 0.01$). Bonferroni correction was applied to the Kellis and Wolfe data separately.

	Enrichment Test			Simulation Test		
Breakpoints (Kellis or Wolfe Scer-Kwal) within 1 kb of a Genomic Feature	Total number of bins with breakpoint	Number of bins with both an Scer-Kwal and an experimental break	*P*-value	Observed ≤ 1kb	Mean	*P*-value
All Kellis *S. cerevisiae – K. waltii* breaks	902	79	0.3477	149	125	0.0303
Kellis breakpoints : tRNAs	124	31	$<10^{-8}$	74	23.5	$<10^{-4}$
Kellis breakpoints : LTRs	98	31	$<10^{-10}$	69	19.6	0.0001
Kellis breakpoints : hcARSs	183	23	0.0295	62	32.1	0.0006
Kellis breakpoints: McCune Rad53 unregulated origins (early firing)	52	10	0.0098	26	9.8	0.0031
All Wolfe *S. cerevisiae – K. waltii* breaks	621	53	0.4810	106	91.2	0.1054
Wolfe breakpoints : tRNAs	106	28	$<10^{-7}$	62	20.5	$<10^{-4}$
Wolfe breakpoints : LTRs	76	25	$<10^{-9}$	55	15.7	0.0001
Wolfe breakpoints : hcARSs	125	18	0.0149	37	22.1	0.0246
Wolfe breakpoints : McCune Rad53 unregulated origins (early firing)	42	10	0.0019	24	8.1	0.0032

Table S13: Enrichment analysis among various genomic features. Significant *P*-values ($P < 0.05$) are highlighted in yellow, and the entire box is highlighted in pink for those significant after a Bonferroni correction ($P < 0.0004$).

	Telomeres	Telomeric repeats	tRNAs	Tys	LTRs	Centromere	Ribosomal protein genes	snoRNAs	snRNAs	Spo11 hotspots
Spo11 hotspots	1.000	1.000	0.590	0.999	0.996	0.949	0.963	0.706	0.604	NA
tRNAs	1.000	0.997	NA	2.41E-10	5.09E-125	1.000	0.989	0.613	1.000	0.590
Tys	0.583	0.754	2.41E-10	NA	1.30E-51	1.000	0.992	1.000	1.000	0.999
LTRs	2.43E-04	2.43E-04	5.09E-125	1.30E-51	NA	1.000	1.000	0.617	0.427	0.996
hcARSs	0.264	0.014	1.22E-07	0.968	6.21E-08	0.107	0.919	0.396	1.000	0.979
OriDB ARSs	8.98E-10	1.55E-15	0.003	0.520	3.72E-07	0.302	0.955	0.865	1.000	0.948
Confirmed	0.006	4.61E-04	5.65E-04	0.763	2.01E-05	0.145	0.914	0.815	1.000	0.974
Likely	1.46E-04	1.94E-08	0.304	0.901	0.009	0.483	0.733	0.728	1.000	0.826
Dubious	0.004	0.052	0.256	0.028	0.031	1.000	0.900	0.781	1.000	0.255
McCune origins	1.000	1.000	1.61E-07	0.915	5.34E-05	0.386	0.847	0.560	1.000	0.798
nonCDR (early firing)	1.000	1.000	5.83E-06	1.000	7.24E-04	0.090	0.976	0.859	1.000	0.923
CDR (late firing)	1.000	1.000	0.003	0.651	0.015	1.000	0.500	0.363	1.000	0.516
Rad53 unregulated (early firing)	1.000	1.000	3.81E-08	0.515	6.14E-07	0.141	0.999	0.723	1.000	0.832
Rad53 regulated (late firing)	1.000	1.000	0.088	1.000	0.473	1.000	0.152	0.447	1.000	0.620

Table S14: Regression analysis for the evolutionary breakpoints. Logistic regression analysis produced the following reduced model: logit(breakpoint) = c_{tRNA}(tRNA) + c_{hcARS}(hcARS) + $c_{telomeric\ repeat}$(telomeric repeat) + $c_{telomere}$(telomere) + $c_{intercept}$. The logit is ln(odds ratio of the probability of breakage), where the probability of breakage is taken from a logistic distribution. The terms tRNA, hcARS, telomeric repeat, and telomere are a value (0, 1, 2, etc.) referring to the number of occurrences of that feature in a given 5 kb bin in the genome. [1]Coefficients in the model reflecting the effect of the genomic feature on the likelihood of a bin having a breakpoint. [2]The 95% confidence interval as determined by 1,000 bootstrap simulations. [3]The *P*-value of that term in the model. The Wolfe regression model does not find the telomere term to be significant. In both datasets, the most telomeric element negatively impacts breakpoint prediction. This negative correlation is seen because the most telomeric-proximal regions are bound by genes on only one side and thus by definition are not intergenic and are unable to bear breakpoints. The Wolfe model lacks the telomere term perhaps as a result of the Wolfe *S. cerevisiae – K. waltii* homology set excluding more telomeric genes than the Kellis dataset.

Term	Coefficient[1]	95% bootstrap confidence interval[2]	*P*-value in model[3]
Kellis *S. cerevisiae – K. waltii* breakpoints			
Intercept	-0.6396	[-0.7389, -0.5528]	$< 10^{-15}$
Telomeric repeats	0.7497	[0.3539, 1.259]	0.0029
Telomeres	-2.6404	[-4.760, -1.390]	0.0009
hcARSs	0.3839	[0.1781, 0.5936]	0.0004
tRNAs	0.4123	[0.1863, 0.6585]	0.0007
Wolfe *S. cerevisiae – K. waltii* breakpoints			
Intercept	-1.1724	[-1.2818, -1.0640]	$< 10^{-15}$
Telomeric repeats	-0.8004	[-14.074, -0.16106]	0.0995
Telomeres	-13.149	[-14.372, -12.086]	0.9572
hcARSs	0.2704	[0.02965, 0.48843]	0.0213
tRNAs	0.6416	[0.39997, 0.90655]	$< 10^{-6}$

References

Admire A, et al. 2006. Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. Genes Dev. 20:159-173.

Aladjem MI, Koltin Y and Lavi S. 1988. Enhancement of copper resistance and CupI amplification in carcinogen-treated yeast cells. Mol Gen Genet. 211:88-94.

Argueso JL, et al. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. Proc Natl Acad Sci U S A. 105:11845-11850.

Bond U, Neal C, Donnelly D and James TC. 2004. Aneuploidy and copy number breakpoints in the genome of lager yeasts mapped by microarray hybridisation. Curr Genet. 45:360-370.

Borde V, et al. 2004. Association of Mre11p with double-strand break sites during yeast meiosis. Mol Cell. 13:389-401.

Brown CJ, Todd KM and Rosenzweig RF. 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol. 15:931-942.

Byrne KP and Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 15:1456-1461.

Chen C and Kolodner RD. 1999. Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants. Nat Genet. 23:81-85.

Degtyareva NP, Chen L, Mieczkowski P, Petes TD and Doetsch PW. 2008. Chronic oxidative DNA damage due to DNA repair defects causes chromosomal instability in Saccharomyces cerevisiae. Mol Cell Biol. 28:5432-5445.

Dunham MJ, et al. 2002. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 99:16144-16149.

Dunn B, Levine RP and Sherlock G. 2005. Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC Genomics. 6:53.

Fischer G, James SA, Roberts IN, Oliver SG and Louis EJ. 2000. Chromosomal evolution in Saccharomyces. Nature. 405:451-454.

Gordon JL, Byrne KP and Wolfe KH. 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern Saccharomyces cerevisiae genome. PLoS Genet. 5:e1000485.

Gresham D, et al. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 4:e1000303.

Hall C, Brachat S and Dietrich FS. 2005. Contribution of horizontal gene transfer to the evolution of Saccharomyces cerevisiae. Eukaryot Cell. 4:1102-1115.

Huang D and Koshland D. 2003. Chromosome integrity in Saccharomyces cerevisiae: the interplay of DNA replication initiation factors, elongation factors, and origins. Genes Dev. 17:1741-1754.

Hwang JY, et al. 2008. Smc5-Smc6 complex suppresses gross chromosomal rearrangements mediated by break-induced replications. DNA Repair (Amst). 7:1426-1436.

James TC, Usher J, Campbell S and Bond U. 2008. Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. Curr Genet. 53:139-152.

Kellis M, Birren BW and Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature. 428:617-624.

Kellis M, Patterson N, Endrizzi M, Birren B and Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature. 423:241-254.

Koszul R, Caburet S, Dujon B and Fischer G. 2004. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. Embo J. 23:234-243.

Lemoine FJ, Degtyareva NP, Kokoska RJ and Petes TD. 2008. Reduced levels of DNA polymerase delta induce chromosome fragile site instability in yeast. Mol Cell Biol. 28:5359-5368.

Lemoine FJ, Degtyareva NP, Lobachev K and Petes TD. 2005. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. Cell. 120:587-598.

Maxwell PH and Curcio MJ. 2007. Retrosequence formation restructures the yeast genome. Genes Dev. 21:3308-3318.

McCune HJ, et al. 2008. The temporal program of chromosome replication: genomewide replication in clb5[Delta] Saccharomyces cerevisiae. Genetics. 180:1833-1847.

Nieduszynski CA, Hiraga S, Ak P, Benham CJ and Donaldson AD. 2007. OriDB: a DNA replication origin database. Nucleic Acids Res. 35:D40-46.

Payen C, Koszul R, Dujon B and Fischer G. 2008. Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. PLoS Genet. 4:e1000175.

Perez-Ortin JE, Querol A, Puig S and Barrio E. 2002. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. Genome Res. 12:1533-1539.

Rachidi N, Barre P and Blondin B. 1999. Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of Saccharomyces cerevisiae. Mol Gen Genet. 261:841-850.

Schacherer J, de Montigny J, Welcker A, Souciet JL and Potier S. 2005. Duplication processes in Saccharomyces cerevisiae haploid strains. Nucleic Acids Res. 33:6319-6326.

Schacherer J, Tourrette Y, Potier S, Souciet JL and de Montigny J. 2007. Spontaneous duplications in diploid Saccharomyces cerevisiae cells. DNA Repair (Amst). 6:1441-1452.

Schacherer J, Tourrette Y, Souciet JL, Potier S and De Montigny J. 2004. Recovery of a function involving gene duplication by retroposition in Saccharomyces cerevisiae. Genome Res. 14:1291-1297.

Tourrette Y, et al. 2007. Spontaneous deletions and reciprocal translocations in Saccharomyces cerevisiae: influence of ploidy. Mol Microbiol. 64:382-395.

Umezu K, Hiraoka M, Mori M and Maki H. 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.

VanHulle K, et al. 2007. Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. Mol Cell Biol. 27:2601-2614.

B.

Figure S1

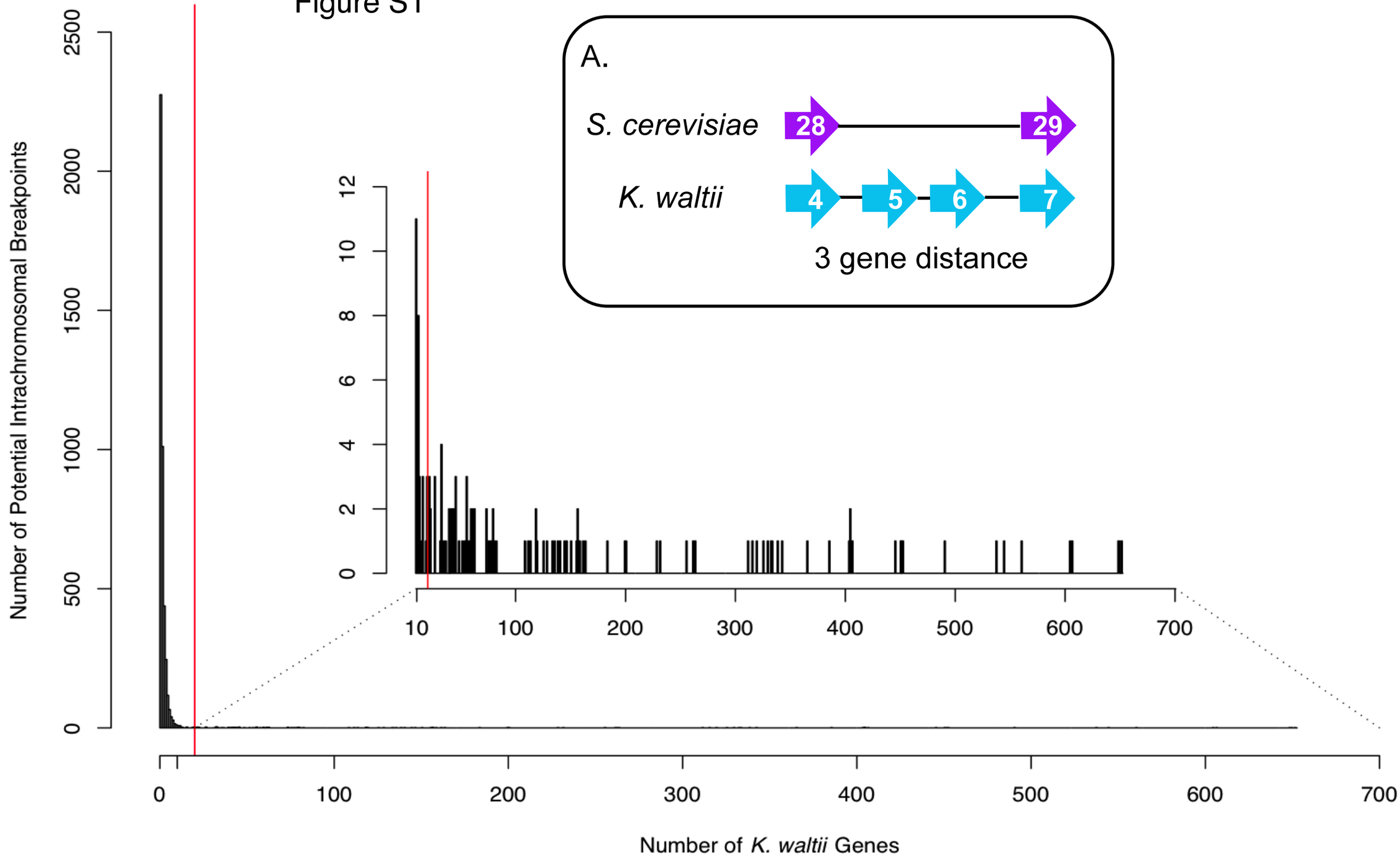

A.
S. cerevisiae
28
29
K. waltii
4
5
6
7
3 gene distance
Number of Potential Intrachromosomal Breakpoints
0
500
1000
1500
2000
2500
0
2
4
6
8
10
12
10
100
200
300
400
500
600
700
0
100
200
300
400
500
600
700
Number of K. waltii Genes

Figure S2

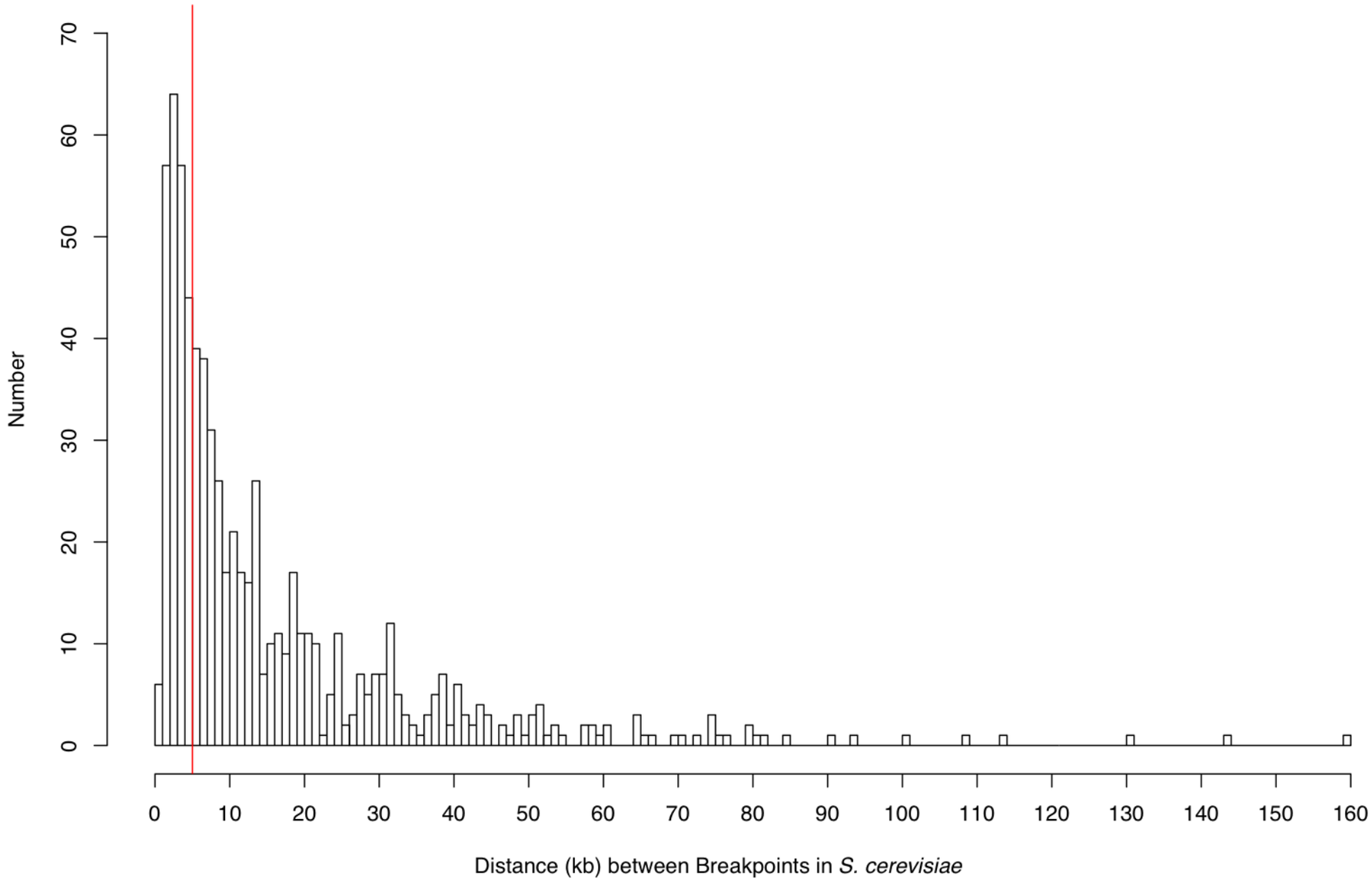
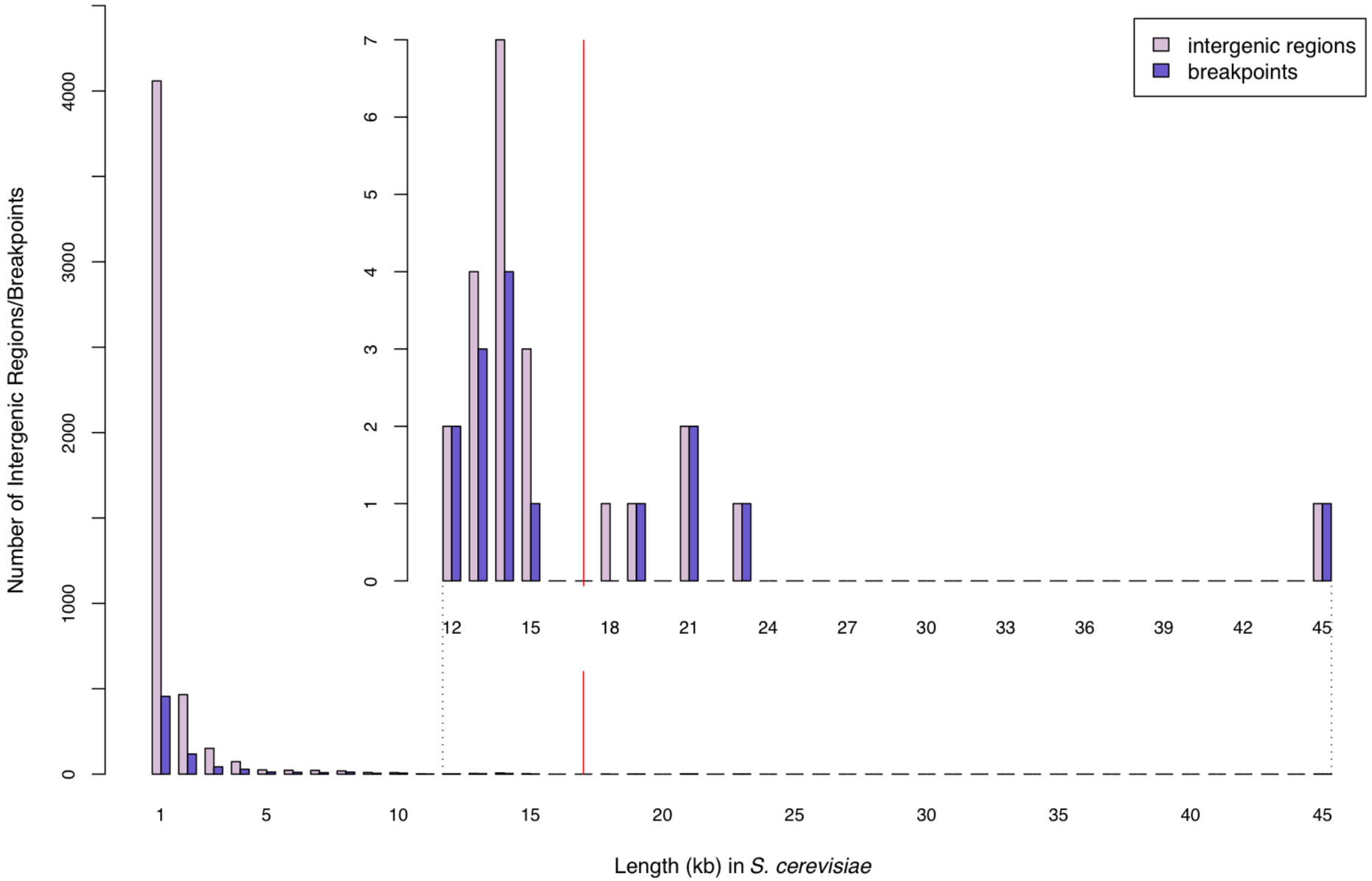

Number
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
Distance (kb) between Breakpoints in S. cerevisiae

Figure S3


intergenic regions
breakpoints
Number of Intergenic Regions/Breakpoints
0
1000
2000
3000
4000
0
1
2
3
4
5
6
7
12
15
18
21
24
27
30
33
36
39
42
45
1
5
10
15
20
25
30
35
40
45
Length (kb) in S. cerevisiae