

Supporting Information

Serero et al. 10.1073/pnas.1314423111

SI Materials and Methods

Yeast Strains and Constructions. All strains used in this study are derivatives of BY4741 (*MATa his3ΔI leu2Δ0 met15Δ0 ura3Δ0*) (1). Gene deletions were made by PCR-based gene disruption methods using genomic DNA from the collection of the *Saccharomyces cerevisiae* single-gene deletants (2). The PCR-amplified DNA fragment, which contains the KanMX4-selectable marker flanked by the upstream and downstream sequence of each ORF of interest, was transformed into BY4741, and the gene deletions were verified by PCR and Southern blot. The double *cac1Δ cac3Δ* and triple *mec1Δ tel1Δ sml1Δ* mutant haploid were obtained by single mutant mating-type switching, crossing, sporulation, and isolation of the appropriate strains, ultimately verified by PCR. The *sml1Δ* mutation is necessary to suppress the lethality of the *mec1Δ* mutation (3); in the main text, the *mec1Δ tel1Δ sml1Δ* genotype is abbreviated to *mec1Δ tel1Δ*. Introduction of the gross chromosomal rearrangement (GCR) assay in the WT (BY4741) and derivative *pif1Δ* mutant was performed by replacing *HXT13* (~7.5 kb telomeric to *CAN1*) with a *URA3* cassette.

Production of the Mutation Accumulation Lines. In most instances (Fig. 1), four parallel mutation accumulation (MA) lines were derived from each parental strain and subjected to 75 (*mre11Δ* and *cac1Δ cac3Δ* mutants) or 100 (WT, *rad27Δ*, *pif1Δ*, *msh2Δ*, *tsa1Δ*, *pol32Δ*, and *clb5Δ* mutants) single-cell bottlenecks, except for the *mec1Δ tel1Δ sml1Δ* mutant whose eight derived lines were passed through 25 bottlenecks. The single-cell bottleneck was performed by picking a random average-sized colony and by streaking it for individual colony forming units on rich yeast extract/peptone/dextrose (YPD) plates for 2–4 d of growth at 30 °C.

Mutation Assays. The canavanine resistance (Can^r) and GCR assays were performed as described in refs. 4 and 5, respectively.

Next-Generation Sequencing. Genomic DNA from 44 yeast strains (four lines per strain but eight for *mec1Δ tel1Δ sml1Δ*) was sequenced using V4 and 5500 SOLiD instruments (Life Technologies) following the manufacturer's standard protocols (Life Technologies, v.4 protocol). Libraries were constructed for (50 + 35 bp) paired-end sequencing (with an insert size of ~145 bp), barcoded, and mixed in four pools: WT, *pif1Δ* and *rad27Δ* (pool 1), *msh2Δ*, *tsa1Δ*, *pol32Δ* and *clb5Δ* (pool 2), *mre11Δ* and *cac1Δ cac3Δ* (pool 3), and *mec1Δ tel1Δ sml1Δ* (pool 4). On average, 8 × 10⁷ reads per line were obtained.

Reads Mapping. Mapping of the read was performed by programs provided by Life Technologies. For SNP calling in the unique regions, small InDel calling, and calling of intermediate-sized and large structural variants, nonredundant reads were aligned with the BioScope (v.1.3) and LifeScope (2.5) (Life Technologies). For SNP calling in the multialigned (M) regions and identification of structural variations, the reads were aligned with mapreads from corona-lite4.2.2 (Life Technologies). Intragenomic repetitive DNA such as Ty elements and multigene families affect the accuracy of the read alignment and are a source of false-positive mutation calls. Before mutation analysis, we identified the uniquely (U) and multialigned (M) regions across the reference genome of *S. cerevisiae* by analyzing multialignments from high-throughput sequencing simulation (HTS simulation). The coverage calculation was performed on the U regions. After removal of the PCR duplicates (average, 25%), it varies between ~75× and 153× (average, ~105×). The reads were mapped on the

S288c reference genome, version June 2008 (S288C_reference_genome_R61-1-1_20080605, SGD *Saccharomyces* Genome Database; www.yeastgenome.org/). Then, the LiftOver Genome Annotations tool (6) (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) was used to convert mutation coordinates from the S288C genome version of June 2008 to the version of February 2011 (S288C_reference_genome_R64-1-1_20110203, SGD *Saccharomyces* Genome Database; www.yeastgenome.org/).

Bioinformatics Methods of Mutation Calling and Molecular Validations. An extensive set of bioinformatics pipelines, diagrammed in Fig. S1, was used to detect all mutation types. Sequence variations recurrently found in the MA lines, thus corresponding to constitutive variations between the reference S288C genome and the BY4741 strain background, were discarded.

Base substitution (SNP) calling in the U regions was performed with BioScope (v1.3). In addition to default parameters, the following rules were applied: (i) calls should be detected on both DNA strands, (ii) only unique reads were used for SNP calling, (iii) adjacent SNPs should be detected, (iv) the values for Reads Max Mismatch Alignmentlength and Reads Min Alignmentlength Readlength ratios were set to 0.5, and (v) high stringency criteria were applied. Finally, generated calls were considered robust if they were reported in at least 70% of the reads covering the SNP position. Differently, SNP calling in the M regions, was performed by using adjacent unique sequence variations defined by the HTS simulation. SNP calls were made by analysis of base changes in 50-bp read alignments with AnnotateChanges (v0.2, corona-1.0.1r0-4) provided by ABI (Life Technologies). A call was selected if the base change was consistent with a minimal of 10 read alignments on each strand with a minimal of five different alignment starting points. Within the M regions, calls within homopolymers and microsatellites tracts remained difficult to ascertain and therefore were eliminated.

Small InDel calls (–11 to +3 bp) were performed in the U regions with the BioScope v1.3 software using the default values. Calls supported by at least six nonredundant reads and by at least three reads on each strand were retained.

Annotation of the SNP and small InDel mutations were determined by the Variant Effect Predictor tool (7) (www.ensembl.org/tools.html).

The various structural variants (SVs) were detected in the U regions from inconsistencies in paired-end read alignments identified with SVDetect (8). The coordinates of the large InDels were found by identifying significant changes in coverage along chromosomes with FREEC (9). Only unique alignments were used for FREEC analysis. Copy number variation was calculated by dividing the mean coverage of the large InDel region by the average of haploid genome coverage. Intermediate-sized and large SV calls (>50 bp) were detected with SVDetect. The size of the sliding window used for partitioning the yeast genome was defined as $\mu + 2$ SD, with μ and SD corresponding to the values of mean insert size and of the SD (both values were determined by SVDetect), respectively. Our 44 paired-end libraries had insert size varying between 116 and 188 bp. The size of the overlapping windows was fixed to be equal to half of the sliding window size. The minimal requirement for a mutational call was 20 paired-end reads.

The aneuploidies were detected by comparing the coverage of each chromosome to the coverage of the chromosome X unique regions because this chromosome did not exhibit copy number variations across all of the MA lines.

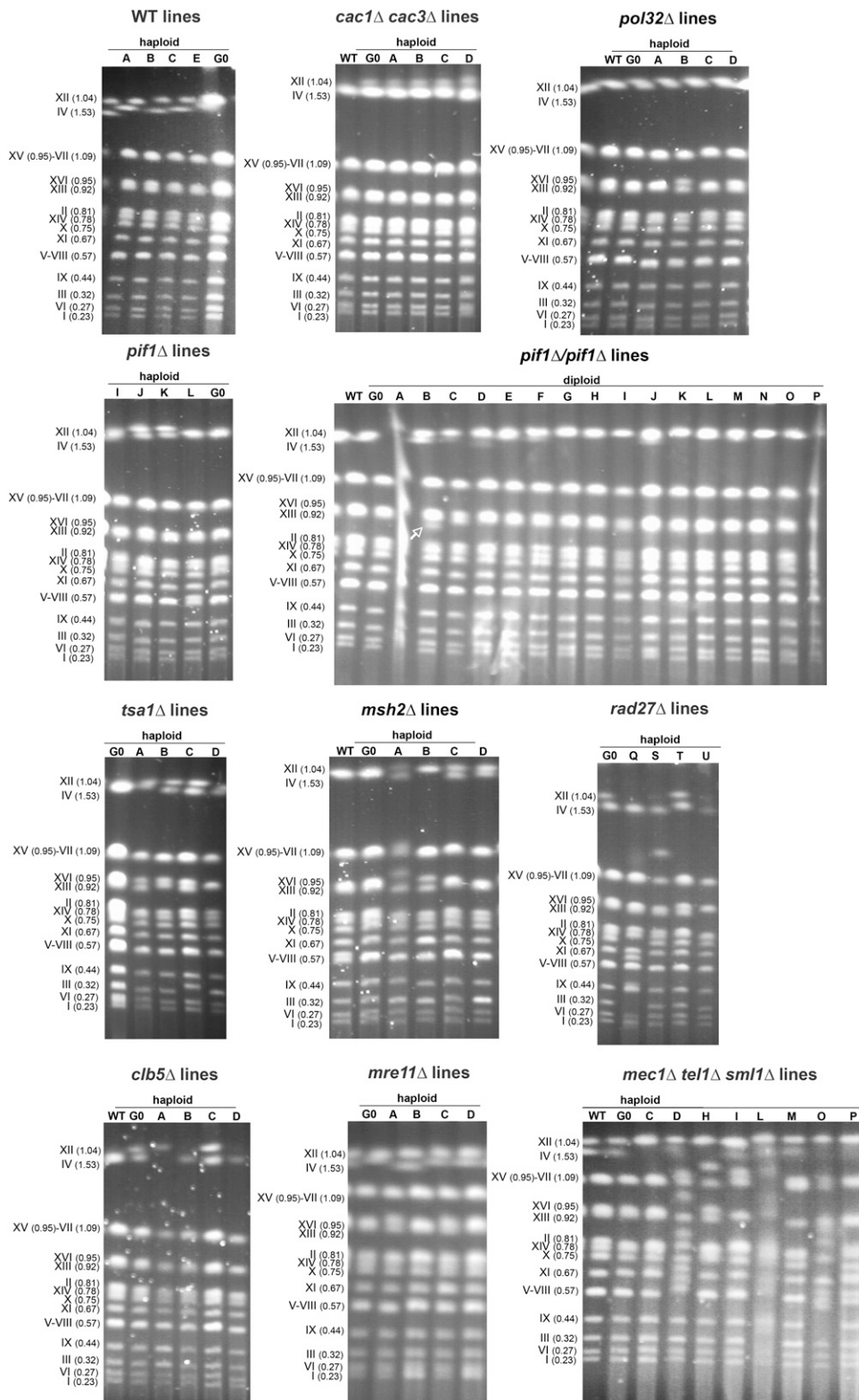


Fig. S2. PFGE analyses of haploid MA lines. For each mutant, a set of four haploid lines was subjected to 100 bottleneck passages, except the *cac1Δ cac3Δ* and *mre11Δ* lines (75 bottleneck passages). The eight haploid *mec1Δ tel1Δ sml1Δ* lines were passaged until 25 bottlenecks. For the *pif1Δ* mutant, an additional set of 16 homozygous diploid parallel lines were subjected to 25 bottlenecks. The letter corresponding to each line is given on the top of the panels; the mutant parental strain is indicated by G0. WT corresponds to the haploid BY4741 strain. Chromosomes indicated with arrows are discussed in [Datasets S1–S10](#). Chromosome numbers and their expected sizes (Mb) are given on the left.

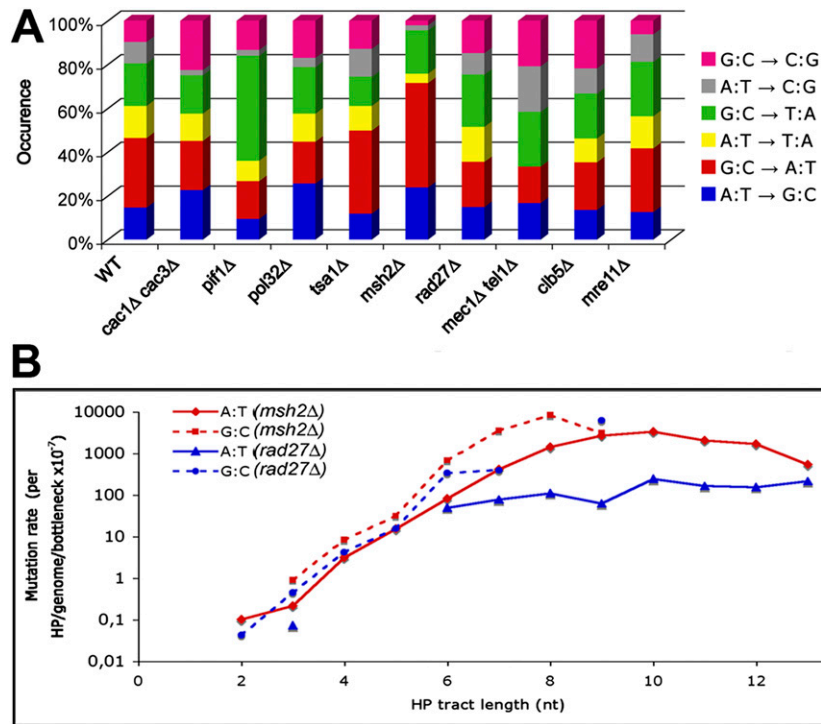


Fig. 53. Mutation profiles. (A) Base substitution profiles of the WT and mutant lines. The total number of mutations per strain is shown on the top. (B) The rate of +1/-1 InDels mutagenesis in the *msh2Δ* and *rad27Δ* lines depends on homopolymer tract length. The mutation rate was calculated on the basis of the following formula: (number of dN mutations)/(number of dN runs in the genome)/(100 bottlenecks). We determined the number of dN runs of a given length in unique regions of the genome and then calculated the number of mutations occurring in these dN runs in the four *msh2Δ* (red line) and *rad27Δ* (blue line) mutants. This calculation was performed for poly(A:T) and poly(G:C) tracts up to 14 nucleotides long. Surprisingly, no +1/-1 InDels located in poly(A) and poly(T) tracts between 4 and 5 bp long and in poly(G) and poly(C) tracts of 8 bp long were identified in the *rad27Δ* lines for a reason that remains to be elucidated.

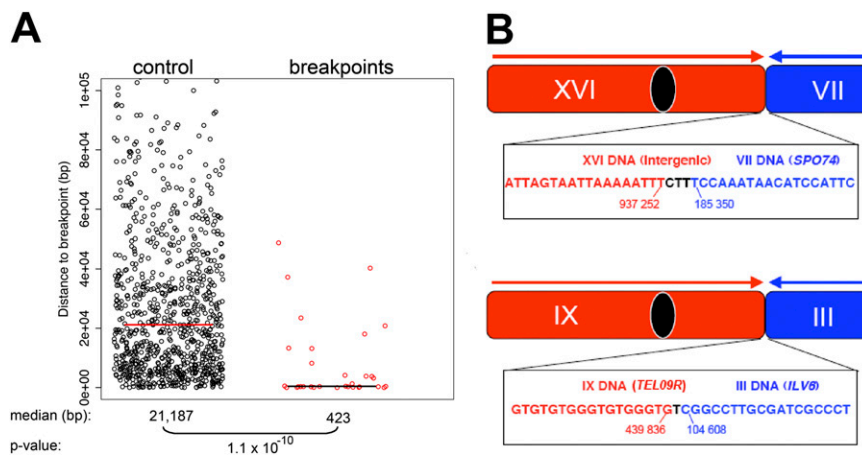


Fig. 54. Profile of chromosome rearrangement breakpoints. (A) Enrichment of chromosome breakpoints in the vicinity of the Ty elements. The median distance (bp) between the unique breakpoints identified in the mutant lines ($n = 33$) (*tsa1Δ*, *msh2Δ*, *rad27Δ*, *clb5Δ*, *mre11Δ* and *mec1Δ*, *tel1Δ* *sm11Δ*) and the nearest Ty and/or LTR elements was calculated. As a control, the same analysis was performed between a set of random coordinates in unique regions ($n = 918$) and the nearest Ty and/or LTR elements. The sets of random coordinates and breakpoints are represented in black and red rhombus, respectively. The P value was calculated with a Wilcoxon test of the median, two-tailed. (B) Breakpoint junction sequences of two nonreciprocal translocations identified in the *mec1Δ* *tel1Δ* *sm11Δ* lines. Panels show the structure of the translocations identified in the M (Top) and O lines (Bottom). The sequence identity between the two chromosomes at the junction is colored in black. As the sequence junctions share low homology (only up to three nucleotides), these chromosome rearrangements are likely the product of nonhomologous end joining events. The name of the gene is indicated in parentheses when the breakpoint is intergenic. *TEL09R* corresponds to the telomeric region on the right arm of chromosome IX.

Table S1. Rate of Can^r progeny in MA lines (G₀ and G₁₀₀ bottlenecks)

Strain	Mutation accumulation lines*	Can ^r mutation rate (×10 ⁻⁷)	95% CI [†]		Fold increase [‡]
			Upper	Lower	
WT	G ₀	2.80	0.77	2.20	1.0
	G ₁₀₀ -A	2.35	0.50	1.50	0.8
	G ₁₀₀ -B	1.62	0.37	0.77	0.6
	G ₁₀₀ -C	2.49	1.09	0.73	0.9
	G ₁₀₀ -E	3.05	0.71	1.70	1.1
<i>cac1Δ cac3Δ</i>	G ₀	1.16	0.23	1.02	1.0
	G ₇₅ -A	0.82	0.26	0.77	0.7
	G ₇₅ -B	0.97	0.43	1.03	0.8
	G ₇₅ -C	0.48	0.19	0.93	0.4
	G ₇₅ -D	0.48	0.18	0.93	0.4
<i>pif1Δ</i>	G ₀	20.41	7.10	8.42	1.0
	G ₁₀₀ -I	16.83	7.83	9.02	0.8
	G ₁₀₀ -J	13.38	6.77	5.72	0.7
	G ₁₀₀ -K	11.21	4.50	6.29	0.5
	G ₁₀₀ -L	7.31	3.12	4.26	0.4
<i>pol32Δ</i>	G ₀	2.81	1.04	2.44	1.0
	G ₁₀₀ -A	2.07	0.59	0.70	0.7
	G ₁₀₀ -B	1.37	0.33	0.80	0.5
	G ₁₀₀ -C	2.81	0.99	1.44	1.0
	G ₁₀₀ -D	2.11	0.70	2.46	0.7
<i>tsa1Δ</i>	G ₀	19.25	3.79	8.23	1.0
	G ₁₀₀ -A	22.42	9.95	5.61	1.2
	G ₁₀₀ -B	9.31	2.38	5.70	0.5
	G ₁₀₀ -C	17.00	6.99	7.64	0.9
	G ₁₀₀ -D	7.25	1.86	6.49	0.4
<i>msh2Δ</i>	G ₀	59.60	38.19	88.17	1.0
	G ₁₀₀ -A [§]	ND	—	—	—
	G ₁₀₀ -B	20.21	5.40	5.97	0.3
	G ₁₀₀ -C	34.97	9.57	13.99	0.6
	G ₁₀₀ -D	66.76	27.49	3.95	1.1
<i>rad27Δ</i>	G ₀	95.87	24.32	101.07	1.0
	G ₁₀₀ -Q [§]	ND	—	—	—
	G ₁₀₀ -S	126.19	57.92	160.24	1.3
	G ₁₀₀ -T	90.74	45.20	46.43	0.9
	G ₁₀₀ -U	17.91	26.02	6.30	0.2
	G ₀	482.48	66.34	294.10	—
<i>mec1Δ tel1Δ sml1Δ</i>	G ₀	482.48	66.34	294.10	—
	<i>clb5Δ</i> G ₀	2.28	0.85	4.72	1.0
	G ₁₀₀ -A	4.19	1.12	3.73	1.8
	G ₁₀₀ -B	1.95	0.60	1.67	0.9
	G ₁₀₀ -C [§]	ND	—	—	—
<i>mre11Δ</i>	G ₁₀₀ -D	2.40	0.72	2.25	1.1
	G ₀	20.81	9.10	4.18	1.0
	G ₇₅ -A	24.93	6.22	4.02	1.2
	G ₇₅ -B	23.58	10.52	7.76	1.1
	G ₇₅ -C	19.73	7.43	7.30	0.9
	G ₇₅ -D	3.15	1.41	3.35	0.2

The CAN^r assay was performed as previously described (4). Strains are as in Table 2. ND, not determined.

*G₀ or G₁₀₀ indicated the number of bottleneck passages, followed by the clone line (letters).

[†]95% CIs of the median were calculated by Luria-Delbruck fluctuation analysis (Lea Coulson Median).

[‡]Fold increase relative to the parental (G₀) rate.

[§]The CAN^r assay could not be performed because the cell line became diploid (*rad27Δ* G₁₀₀-Q) (Fig. 3), acquired an arginine deficiency (*msh2Δ* G₁₀₀-A) (Table S2), or became disomic for chromosome V that carries the *CAN1* gene (*clb5Δ* G₁₀₀-C) (Table S4).

Table S2. Phenotypic growth analysis of the mutation accumulation lines after bottleneck cycles

Strain	Mutation accumulation line*	Growth on YPD [†]			Growth on YPG [‡]	Growth on synthetic supplemented media [§]	Mating efficiency [¶]
		23 °C	30 °C	37 °C			
WT	G ₁₀₀ -A	+++	+++	+++	+++	+++	+++
	G ₁₀₀ -B	+++	+++	+++	+++	+++	+++
	G ₁₀₀ -C	+++	+++	+++	+++	+++	+++
	G ₁₀₀ -E	+++	+++	+++	+++	+++	+++
<i>cac1Δ cac3Δ</i>	G ₇₅ -A	+++	+++	+++	+++	+++	+++
	G ₇₅ -B	+++	+++	+++	+++	+++	+++
	G ₇₅ -C	+++	+++	+++	+++	+++	+++
	G ₇₅ -D	+++	+++	+++	+++	+++	+++
<i>pif1Δ</i>	G ₁₀₀ -I	+++	+++	+++	–	+++	+++
	G ₁₀₀ -J	+++	+++	+++	–	+++	+++
	G ₁₀₀ -K	+++	+++	+++	–	+++	+++
<i>pol32Δ</i>	G ₁₀₀ -L	+++	+++	+++	–	+++	–
	G ₁₀₀ -A	+++	+++	+++	+++	+++	+++
	G ₁₀₀ -B	+++	+++	+++	+++	+++	+++
	G ₁₀₀ -C	+++	+++	+++	+++	+++	+++
<i>tsa1Δ</i>	G ₁₀₀ -D	+++	+++	+++	+++	+++	+++
	G ₁₀₀ -A	+++	+++	+++	+++	+++	–
	G ₁₀₀ -B	–	+++	+++	–	+	+++
	G ₁₀₀ -C	–	+++	–	+++	+++	–
<i>msh2Δ</i>	G ₁₀₀ -D	+++	+++	–	+++	+++	–
	G ₁₀₀ -A	+	+	–	–	+, arg-	–
	G ₁₀₀ -B	+++	+	+++	–	+++, trp-	+++
	G ₁₀₀ -C	+++	+	–	–	+++	–
<i>rad27Δ</i>	G ₁₀₀ -D	+++	+	–	–	+	–
	G ₁₀₀ -Q	+++	+++	–	–	+++, arg-,lys-	–
	G ₁₀₀ -S	+++	+++	–	–	+++	+++
	G ₁₀₀ -T	+++	+++	+++	+	+++	+++
<i>mec1Δ tel1Δ sml1Δ</i>	G ₁₀₀ -U	+++	+++	–	–	+++	–
	G ₂₅ -C	+++	+++	+++	–	+	+++
	G ₂₅ -D	+	+	+	–	+	+++
	G ₂₅ -H	+	+	–	–	+	+++
	G ₂₅ -I	+	+	–	–	+	+++
	G ₂₅ -L	+	+	+	–	+	+++
	G ₂₅ -M	+	+	+	–	+	+++
	G ₂₅ -O	+	+	+	–	+	+++
	G ₂₅ -P	+	+	+	–	+	+++
<i>clb5Δ</i>	G ₇₅ -A	+	+	+	+	+	+++
	G ₇₅ -B	+++	+++	+++	+++	+++	+
	G ₇₅ -C	+++	+++	+++	+++	+++	+++
	G ₇₅ -D	+++	+++	+++	+++	+++	+++
<i>mre11Δ</i>	G ₇₅ -A	+++	+++	+++	+++	+++	+++
	G ₇₅ -B	+++	+++	+++	+++	+++	+++
	G ₇₅ -C	+++	+++	+++	+	+++	+++
	G ₇₅ -D	+++	+++	+++	+	+++	–

+++ , +, and – correspond to normal growth (WT level), reduced growth, and no growth on the corresponding medium, respectively.

*The letter and the number correspond to the name of the line and to the number of passages it went through, respectively.

[†]1% yeast extract, 2% (wt/vol) peptone, 2% (wt/vol) D-glucose.

[‡]1% yeast extract, 2% (wt/vol) peptone, 3% (wt/vol) glycerol.

[§]0,17% YNB w/o AA and AS, 0,5% ammonium sulfate, 2% (wt/vol) D-glucose, dropout powder as required. Growth was tested on the following media: SC-adenine, SC-arginine, SC-histidine, SC-leucine, SC-tryptophan, SC-uracil, SC-lysine, and SC-methionine.

[¶]Qualitatively assayed by crossing with testers strains (*MATa his1* and *MATα his1* strains).

Table S3. Location and nature of the minisatellite size variations observed in the MA lines

Minisatellite	Localization	Number of motifs	Motif size (bp)	MA lines			
				WT	<i>rad27Δ</i>	<i>msh2Δ</i>	<i>tsa1Δ</i>
CYC8	Chr II 463757–463935	11	18	—	G ₁₀₀ -U: expansion (gain of 6 motifs)		
NUM1	Chr IV 757379–759782	13	192	—	G ₁₀₀ -S: contraction (loss of 5–6 motifs)	G ₁₀₀ -D: contraction (loss of 3 motifs)	G ₁₀₀ -B: contraction (loss of 4 motifs)
HKR1	Chr IV 1307331–1308128	19	42	—	G ₁₀₀ -U: contraction (loss of 9–10 motifs)		
				—	G ₁₀₀ -Q: contraction (loss of 4–5 motifs)	G ₁₀₀ -D: expansion (gain of 6 motifs)	G ₁₀₀ -A: contraction (loss of 9–10 motifs)
					G ₁₀₀ -S: contraction (loss of 4–5 motifs)		G ₁₀₀ -D: contraction (loss of 6–7 motifs)
					G ₁₀₀ -T: contraction (loss of 4–5 motifs)		
YIL169c	Chr IX 25310–25856	17	42	—	G ₁₀₀ -U: contraction (loss of 2–3 motifs)		
				—	G ₁₀₀ -Q: contraction (loss of 11 motifs)	G ₁₀₀ -B: contraction (loss of 1 motif)	G ₁₀₀ -B: contraction (loss of 10 motifs)
							G ₁₀₀ -D: expansion (gain of 1 motif)
MUC1 (FLO11)	Chr IX 391078–392752	42	36–45	—	G ₁₀₀ -Q: contraction (loss of 16–17 motifs)		G ₁₀₀ -C: contraction (loss of 23–24 motifs)
					G ₁₀₀ -S: contraction (loss of 23–24 motifs)		G ₁₀₀ -D: contraction (loss of 23–24 motifs)
					G ₁₀₀ -T: contraction (loss of 23–24 motifs)		
					G ₁₀₀ -U: contraction (loss of 35–36 motifs)		
YMR317w	Chr XIII 908149–909052	25	36	—	G ₁₀₀ -Q: expansion (gain of 2 motifs)		G ₁₀₀ -A: expansion (gain of 16–17 motifs)
					G ₁₀₀ -S: expansion (gain of 2 motifs)		

Minisatellite size variation was determined by PCR with oligonucleotides designed in the flanking regions of the minisatellite.

