Supplementary information:
The genome-wide rate and spectrum of spontaneous mutations differs between haploid and diploid yeast

SI Materials and Methods

Measures of growth rates during MA

To measure rates of cell division during MA, two random colonies were sampled from a rotating subset of MA plates following 22–27 h of growth (mean 25 h), by absorbing each colony in a 1 µl drop of sorbitol with a pipette under a dissecting microscope, which captures the vast majority of cells. Cell concentration in a defined volume was determined using hemocytometer counts, and the rate of cell division on the plate was calculated assuming exponential growth.

Measures of growth rates following MA

218 non-petite MA lines and 182 replicates of ancestral controls (across ploidy and RDH54 treatments) were grown from frozen stocks for 2 d at 30°C in liquid YPAD with shaking and subsequently held at 4°C. These cultures were used to initiate multiple blocks of growth assays to obtain at least 9 replicates per line. Each day for 11 days each 4°C culture was serially diluted into liquid medium to a final dilution factor of 1:121 and volume of 165 µl. The 400 cultures were grown in two BioscreenC plate readers at 30°C for 24 h with continuous low-medium shaking with optical density (OD) readings at 420-560 nM every 15 min. All 400 cultures were diluted from 4°C and grown every day, with culture locations within plates randomized daily across the two plate readers. Because of a shaking malfunction, 400 out of the 4400 replicate measures were excluded from subsequent analyses.

To calculate maximum growth rate for each replicate a spline was fit to the relationship between *log*(OD) and time, for the data between the second OD measurement and 16 h of growth, using the *R* function *loess* with degree = 1 and span = 0.2; maximum growth rate was estimated as the maximum derivative of this spline across 99 equally-spaced time points. Maximum growth rate values were analyzed in a linear mixed model with a fixed effect of initial OD, fixed effects and interactions of MA status (MA line vs. ancestral control), *RDH54* status, and ploidy level, as well as random effects of day, plate, and line ID. Using predicted line values from this model accounting for random effects, relative log growth rate for a given MA line was calculated by subtracting the mean of the predicted control values for that treatment.

Phenotype testing and re-initializing MA lines

Yeast culture from backup plates was frozen in 15% glycerol every 7 transfers. The MA lines were periodically tested for the following phenotypes: mating behaviour, as indicated by the presence or absence of growth on minimal media following mixing with a MATa or MATα haploid tester strain carrying different auxotrophic markers from the MA lines; RDH54 status, as indicated by the presence or absence of growth on media containing the drug G418 (*rdh54Δ* lines carry the *KanMX*

cassette, which confers G418 resistance); and respiratory phenotype to detect *petite* mutants, as indicated by the presence or absence of growth on a non-fermentable carbon source (yeast extract-peptone-glycerol medium).

In a few cases frozen culture was used to replace or duplicate an MA line, as follows. We duplicated three lines upon discovering unexpected mating behaviour—a possible indication of contamination—using the latest frozen version of each line that retained the original mating behaviour. Since sequence data ultimately revealed that changes to mating behavior were the result of spontaneous mutations and not contamination (see main text), both the original and duplicate lines were retained. Three lines found to be petite relatively early in the experiment were each replaced with the latest available non-petite frozen version. For two lines found to be petite later in the experiment additional lines were created using the latest available non-petite frozen version, and the petite versions were retained for genotyping but not growth rate analysis. Cases of lines with a shared history of mutation accumulation will potentially share mutations, and this was accounted for in the sequence analysis (see below).

Flow cytometry

Genome complement following mutation accumulation was assessed by measuring SYTOX Green fluorescence with a BD LSRII flow cytometer using a similar protocol to ref. (1). MA lines were tested in two blocks, along with multiple isolates of each ancestral control in each block, with approximately 50000 cells analyzed per line. Any ambiguous or unusual cases were re-run in a third block. Ploidy level was assessed by visually comparing fluorescence peak locations in MA lines with those of ancestral controls.

Sequence analysis

We sequenced the MA lines as described in the main text. We verified correct sample identification by searching raw read data for the presence of *KanMX*- vs. *RDH54*-specific sequences in the appropriate samples, as well as the expected ratio of *MATa*- to *MAT*α-specific sequences. Reads were mapped to the yeast reference genome (version R64-2-1) with BWA *mem* (2) and duplicate reads removed with Picard Tools (http://broadinstitute.github.io/picard/). Known repetitive regions and mating type genes were excluded from subsequent analyses. Eight random samples (two from each ploidy-by-*RDH54* treatment) were initially genotyped using GATK HaplotypeCaller in GVCF mode, applying the appropriate *ploidy* argument for each sample. These samples were jointly genotyped to identify likely consensus variants. Base quality scores were then recalibrated in all samples using GATK BaseRecalibrator, masking likely consensus variants for this step. Finally, all recalibrated samples were individually genotyped using GATK HaplotypeCaller in GVCF mode, applying the appropriate ploidy argument for each sample, and jointly genotyped using GATK GenotypeGVCFs.

The vast majority of initially called variants occurred in all or nearly all samples, or in only one sample. Given the length of our experiment the probability that more than one or two nucleotide sites would mutate multiple times independently is very low (see below). Variants found in multiple samples were

included in our analyses only when those samples shared an evolutionary history, i.e., where an MA line was derived from a pre-existing MA line. In these cases mutations were assigned arbitrarily to the original MA line.

Sites were excluded from further analysis if >10% of samples had no data or if the average depth across samples was >2× the average median depth for that chromosome. Variant calls for a sample were also excluded if sample depth was <4 in haploids, <8 in diploids, if sample depth was >2× the median depth for that chromosome and sample, if map quality was <50, if strand bias was high ($P < 0.01$), or if a putative variant was found in >1 reads in other samples. In addition, variants in haploids were excluded if >10% of reads supported the reference allele, and variants in diploids were excluded if the binomial probability of the mutant allele frequency was <0.1% given an expected frequency of 0.5 (an expected frequency of 1/3 was used for variants on ancestrally-trisomic chromosomes; for variants on newly-trisomic chromosomes we calculated the *P*-value given both 1/3 and 2/3 expectations and took the maximum). We accounted for these coverage and allele frequency criteria when calculating the effective number of callable sites in each MA line.

We were able to confirm 36/38 nuclear SNMs and 26/28 indels tested by Sanger sequencing, with no evidence for a difference in confirmation rate between ploidy levels (odds ratio 1.14, $P = 1$), or between SNMs and indels (odds ratio 1.38, $P = 1$). Non-confirmed variants were excluded from all analyses.

We treated mitochondria (mt) in all lines as effectively haploid for the purposes of variant calling and assumed that true mutations would be fixed (homoplasmic) within lines. Calling mt as diploid did not result in any additional homoplasmic variants passing filters. See below for results on putatively-heteroplasmic mt mutations.

Cases of aneuploidy, where a chromosome copy is gained or lost, were identified by comparing median coverage on a chromosome with the average median coverage of other chromosomes in that sample, excluding chromosome XI. A chromosome gain (loss) was called when relative coverage was >1.35 (<0.65) (3). These thresholds should be particularly permissive when calling aneuploidy in haploids: the expected ratio for a chromosome gain in a haploid would be 2, but the highest ratio observed was 1.31. In diploids we observed two cases of probable tetrasomy, where the coverage ratio was >1.8; these were each scored as single events.

Cases where multiple SNMs were called within 50 bp of one another were classified as multi-nucleotide mutations (MNMs), and each group was scored as a single MNM event. Similarly, groups of events within 50 bp that included an indel were classified as "complex" mutations, and each group was scored as a single indel event.

We classified genic mutations as synonymous, missense or nonsense based on their effects on protein sequence, incorporating multiple changes in the case of MNM events. Complex indels were classified using the most extreme classification of their constituent changes. To address the expected neutral probability that a genic

SNM would be non-synonymous we simulated 10^5 mutation events using the nucleotide transition probabilities of all observed SNMs.

Changes to mating behaviour

Assays of mating behaviour identified two cases of apparent sterility in *MAT*α haploid MA lines (one *RDH54*+ and one *rdh54*Δ; haploidy was confirmed by flow cytometry), and so we can estimate the spontaneous rate of mutation to haploid sterility as 1.27×10^{-5} per haploid genome per generation (95% CI: 1.54×10^{-6} to 4.59×10^{-5}). We identified the most likely causal variants as nonsense mutations in *STE12* (activates genes involved in mating; see https://www.yeastgenome.org) and *SIR3* (null mutant is pheromone-unresponsive), respectively. The line with the *SIR3* mutation was also found to have an unusually large proportion of doubled cells (based on flow cytometry and particle imaging). We also found one case where a diploid *rdh54*Δ MA line acquired the ability to mate with *MAT*a haploids (diploidy was confirmed by flow cytometry). Sanger sequencing and whole-genome sequence data is consistent with homozygosity for α-specific sequence at the MAT locus in this line. Our estimate of the spontaneous rate at which diploids acquire mating ability is 5.58×10^{-6} per diploid genome per generation (95%CI: 1.41×10^{-7} to 3.11×10^{-5}).

Testing for effects of mutations on subsequent mutation rates

We tested whether aneuploidy in diploid lines influenced point mutation rates, considering combinations of aneuploidy type (ancestral vs. *de novo*, gains and losses vs. gains only), point mutation type (SNM vs. indel), and effect type (a "local" effect on the aneuploid chromosome itself vs. a genome-wide effect on all chromosomes), and accounting for power to detect mutations (callable sites, MA generations, and chromosome copy number in the case of ancestral chromosome XI trisomy). Chromosomes with a *de novo* increase in copy number would be expected to harbour more point mutations simply due to the increase in mutable sites, but we do not know at what time points *de novo* copy number increases occurred, and so we do not account for this increase in sites, which could create a false impression that aneuploid chromosomes have higher point mutation rates. However, generalized linear models (Poisson) of mutations per chromosome across all lines (n = 1824 chromosomes) revealed no evidence of a "local" effect of aneuploidy on point mutation rates per site for any aneuploidy or point mutation type (all $P >$ 0.45). Binomial tests comparing aneuploid and non-aneuploid lines revealed no evidence of a genome-wide effect of any aneuploidy type on any point mutation type (all $P > 0.089$). The only comparison approaching significance ($P = 0.089$) is suggestive of a 9.3% lower SNM rate in lines with ancestral trisomy for chromosome XI (rate [95% CI]: 2.73 [2.50, 2.98] $\times 10^{-10}$) than lines without ancestral trisomy (rate [95% CI]: 3.01 [2.80, 3.25] $\times 10^{-10}$).

In our experiment five MA lines exhibited *de novo* aneuploidy for more than one chromosome, with no lines exhibiting aneuploidy for more than two chromosomes. Using 5000 datasets where aneuploidy events among lines were simulated under a multinomial distribution, we find that five or more lines with

multiple aneuploid chromosomes is likely to occur by chance ($P = 0.77$) and that the absence of any lines with more than two aneuploid chromosomes is also likely to occur by chance ($P = 0.51$). To address whether point mutations affected the subsequent point mutation rate we tested whether the number of mutations was clustered among lines (overdispersed), compared to the Poisson expectation, using a generalized linear model within each ploidy level and the function *dispersiontest* in the *R* package *AER* (https://CRAN.R-project.org/package=AER); we found no evidence for overdispersion considering either all point mutations or restricting the analysis to non-synonymous point mutations (all $P > 0.11$).

Mitochondrial mutations and heteroplasmy

In our primary analysis of we considered only mt variants called as homoplasmic. Using a "diploid" ploidy setting and allowing for heteroplasmy when calling mt mutations resulted in 17 additional variants, compared to the 132 homoplasmic mutations detected (Table S1). All but one of these putative heteroplasmic variants were non-genic. When heteroplasmic events are included, the overall haploid:diploid mt mutation rate ratio is 9.29, and the mt mutation rate is significantly higher in haploids than diploids for both SNMs (binomial test: $P < 10^{-13}$) and indels ($P < 10^{-13}$) with no effect of the *RDH54* deletion (SNMs: $P = 0.63$; indels: $P = 0.06$).

Table S1. Summary of mitochondrial mutations.

		Haploid		Diploid	
		RDH54+	*rdh54Δ*	*RDH54+*	*rdh54Δ*
Homoplasmic mutation counts	SNM	30	24	3	3
	MNM	1	1	0	0
	indel	40	24	3	3
Heteroplasmic mutation counts	SNM	4	2	1	0
	MNM	0	1	0	0
	indel	4	1	4	0
Mutation rate* (10^{-8}/bp/gen.)	Homoplasmic	1.25	0.89	0.08	0.10
	All	1.39	0.96	0.15	0.10

*Including SNM, MNM, and indel events.

Additional analyses of mutation rate variation among chromosomes

The following analyses were conducted to examine whether factors other than centromere location *per se* could contribute to mutation rate heterogeneity among chromosomes. We considered whether diploid SNM rates might be reduced on chromosomes with relatively distal centromeres due to mutation rate reductions on short chromosome arms. However, among chromosomes the ratio of mutations on the short vs. long arm is highly correlated with arm length ratio ($r = 0.83$, $t = 5.67$, $df = 14$, $P < 10^{-4}$). SNM rates are negatively correlated with centromere distance for both long and short arms, with no statistical difference in rank correlation coefficient, although we presumably have reduced power to detect a significant relationship for short arms (long: $r_S = -0.77$, 95% CI = –0.95 to –0.40, $P < 0.001$;

short: r_S = –0.28, 95% CI = –0.69 to 0.25, P = 0.29; slope comparison: t = 0.73, P = 0.47).

Centromere location is not significantly correlated with chromosome length (r = 0.10, P = 0.71), detection power (r = 0.08, P = 0.77), subtelomeric DNA content (r = –0.21, P = 0.44), recombination hotspot density ((4); crossover: r = –0.30, P = 0.26; non-crossover: r = –0.24, P = 0.36), or density of replication origins (r = –0.32, P = 0.23). Both subtelomeric DNA and recombination hotspot density are negatively correlated with mutation rate across chromosomes at both ploidy levels, but the effect of centromere location on the diploid mutation rate remains significant when these additional factors are included.

Analysis of growth rates in relation to mutations

We examined growth rates in relation to the number of mutations in each line by fitting linear mixed models by maximum likelihood to our 4000 growth rate measurements with random effects of day and plate and fixed effects of initial OD, *RDH54* status, ploidy level, the number of nonsynonymous point mutations (n = 1170; set to zero in controls), and genome size relative to the euploid state, which accounts for aneuploidy and large duplications. For these analyses we assume chromosome XII is 825 kb longer than the length of the reference sequence, based on the relative depth of coverage we observed in the rDNA repeat region. We modeled the effects of mutations on diploids in two ways: first using the unweighted count of mutations, which effectively assumes mutations have completely dominant effects, and second, weighting each mutation by its allele frequency (0.5 for heterozygous variants, 1 for homozygous variants, and 1/3 or 2/3 for variants on trisomic chromosomes). We assessed the significance of main effects using likelihood ratio tests; non-significant interactions between main effects were removed sequentially.

We detect a significant negative interaction between diploidy and mutation number using both weighted and unweighted mutation counts (weighted: χ^2 = 13.29, P < 0.001; unweighted: χ^2 = 5.91, P < 0.05). Analyzing ploidy levels separately we find evidence for significant negative effects of mutation number on growth rate in diploids (weighted: χ^2 = 19.19, $P < 10^{-4}$; unweighted: χ^2 = 19.56, $P < 10^{-5}$) but not haploids (χ^2 = 0.32, P = 0.57). As an alternative approach to account for aneuploidy when examining the effects of MA on haploids vs. diploids we considered only those lines where MA lines had the same karyotype as the ancestral control (n = 3542 growth rate measurements). This analysis also reveals a significant negative interaction between MA and diploidy (χ^2 = 7.50, P < 0.01).

Maximum likelihood estimates of LOH and mutation rates

We consider the number of homozygous reference (RR), heterozygous SNM (RS), homozygous SNM (SS), heterozygous indel (RI), and homozygous indel (II) callable sites in each line following MA, assuming that numbers of sites are Poisson-distributed and that covariances are small enough that they can be neglected. These values change each cell division due to mutation per haploid site (μ_{SNM}, μ_{indel}) and LOH (ρ) per diploid site according to the differential equations:

$$\frac{dn_{RR}}{dt} = -(2\mu_{SNM} + 2\mu_{indel})n_{RR} + \frac{\rho}{2}n_{RS} + \frac{\rho}{2}n_{RI}$$

$$\frac{dn_{RS}}{dt} = 2\mu_{SNM}n_{RR} - \rho n_{RS}$$

$$\frac{dn_{SS}}{dt} = \frac{\rho}{2}n_{RS}$$

$$\frac{dn_{RI}}{dt} = 2\mu_{indel}n_{RR} - \rho n_{RI}$$

$$\frac{dn_{II}}{dt} = \frac{\rho}{2}n_{RI}$$

The above neglects reverse mutations, which will be rare given the relatively short duration of the experiment, and assumes that mutant homozygotes will predominantly be generated by LOH over the time course of the experiment (i.e., mutation rates are low relative to LOH rates, $\rho > \mu_{SNM}, \mu_{indel}$).

These equations can be solved for their value at generation t, assuming all sites are initially in the RR state, accounting for the number of MA generations in each line. We found the maximum likelihood values and 95% confidence intervals of μ_{SNM}, μ_{indel} and ρ using the sum of the log-likelihood across lines in Mathematica.

Our estimates of mutation and LOH rates allow us to predict the number of heterozygous sites expected per genome in the absence of selection and drift (purifying selection and drift will tend to reduce heterozygosity). Approximating the expected number of heterozygous sites at generation $t + 1$ by

$$N_{het,\,t+1} = N_{het,\,t}\,(1 - \rho) + 2\mu(N_{tot} - N_{het,\,t}),$$

where ρ is the LOH rate per diploid site and μ is the mutation (SNM and indel) rate per haploid site, at equilibrium $N_{het} = 2\mu N_{tot}/(2\mu + \rho)$. Given our ML estimates of ρ and μ, we would predict ~96 heterozygous sites per 12 Mb genome. Wild yeast populations, as well as some laboratory strains, are found to have hundreds to thousands of heterozygous sites per genome (25, 44), suggesting that in some populations mutation rates are higher, LOH rates are lower, selection acts to maintain heterozygosity, or outcrossing occurs between divergent lineages.

Probability of multiple mutations occurring at the same site

These calculations assume $m = 1899$ SNMs occur at $T \approx 11.3 \times 10^6$ sites (the average number of callable sites per MA line). The probability that all mutations will occur at different sites is

$$\prod_{x=1}^{m} \frac{T - (x - 1)}{T}$$

$$= 0.8526$$

This leaves a 14.75% chance that there would be at least one site with multiple hits. The probability that exactly one site will be hit twice is

$$\sum_{j=2}^{m}\left[\left(\prod_{x=1}^{j-1}\frac{T-(x-1)}{T}\right)\left(\frac{j-1}{T}\right)\left(\prod_{x=j+1}^{m}\frac{T-(x-2)}{T}\right)\right]$$

= 0.1360

The probability that exactly two sites will be hit twice is

$$\sum_{k=3}^{m}\left[\left(\frac{k-2}{T}\right)\left(\prod_{x=k+1}^{m}\frac{T-(x-3)}{T}\right)\sum_{j=2}^{k-1}\left[\left(\prod_{x=1}^{j-1}\frac{T-(x-1)}{T}\right)\left(\frac{j-1}{T}\right)\left(\prod_{x=j+1}^{k-1}\frac{T-(x-2)}{T}\right)\right]\right]$$

= 0.0108

Finally, the chance that more than two sites will be hit twice is
1 – 0.8526 – 0.1360 –0.0108 = 0.0006

Thus, it is most likely (85.3%) that we would not have any double hits. With 13.6% (1.1%) probability there would be one (two) double-hit sites, which would be excluded under our mutation calling procedure. It is very unlikely (0.06%) that we would be missing more than two double-hit sites.

References

1. Gerstein AC, Chun H, Grant A, Otto SP (2006) Genomic convergence toward diploidy in Saccharomyces cerevisiae. *PLoS Genet*. doi:10.1371/journal.

2. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

3. Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *PNAS* 111(22):E2310–E2318.

4. Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM (2008) High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454(7203):479–485.

Supplementary Figures

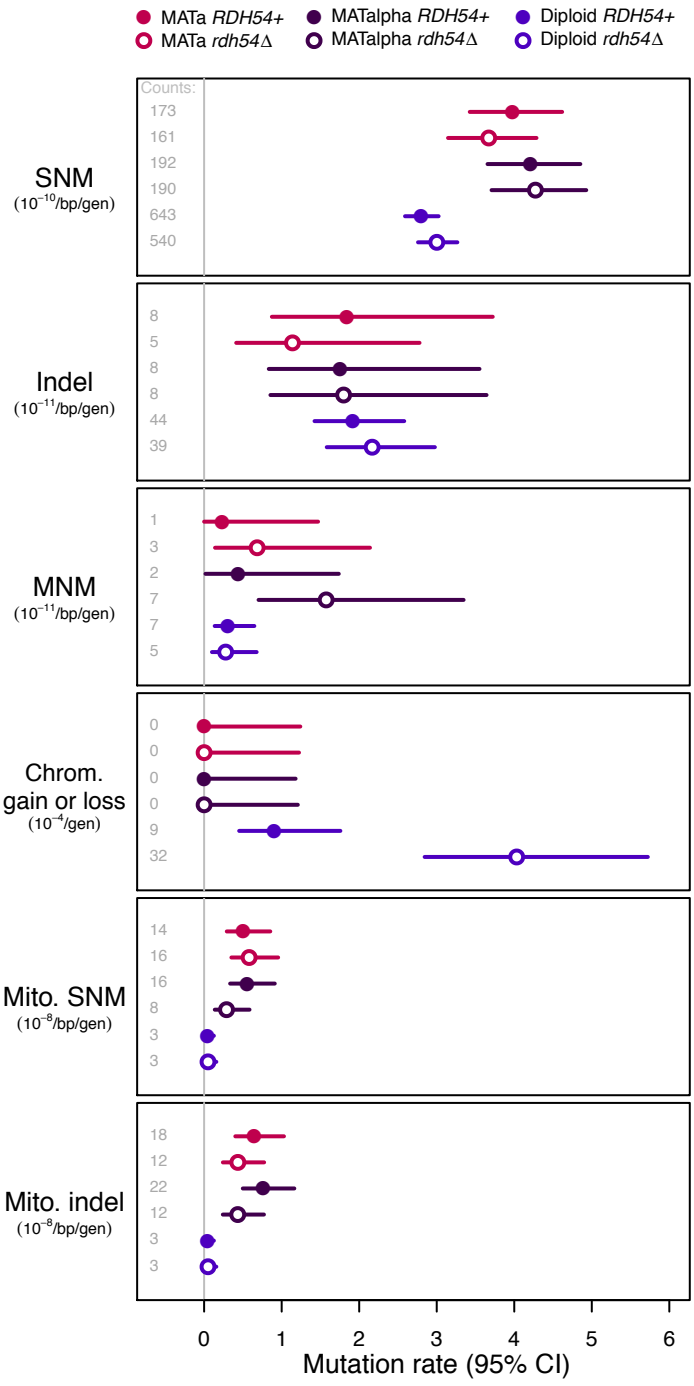

Figure S1. Mutation rates in each group of MA lines, with 95% confidence intervals, showing *MATa* and *MAT*α separately. Mutation categories are as in Fig. 2. Mating type within haploids did not significantly affect the rate of any mutation type.

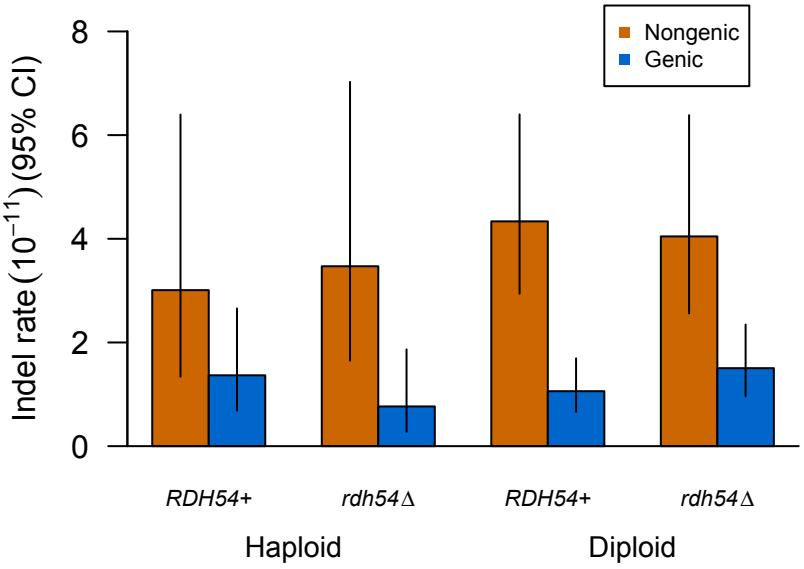

Figure S2. Rates of indel mutation per base pair per generation in each treatment for genic and non-genic regions. Rates of genic indels did not differ among groups (G = 2.56, df = 3, P = 0.47); rates of non-genic indels did not differ among groups (G = 0.93, df = 3, P = 0.82).

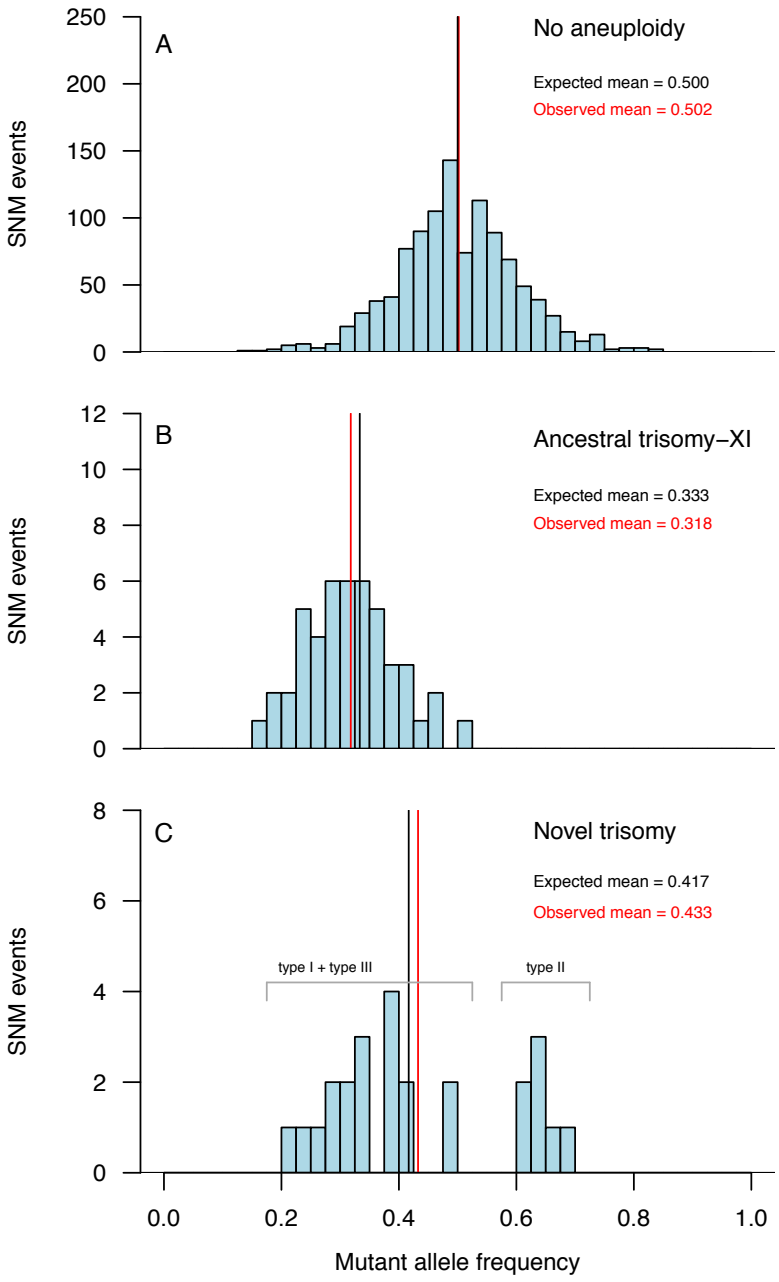

Figure S3. Allele frequencies inferred from depth of coverage for diploid heterozygous SNMs, accounting for different histories of aneuploidy. (A) Most events occurred on non-aneuploid chromosomes, where the observed mean coverage frequency for the mutant allele did not differ from the 0.5 expectation (t = –1.11, P = 0.27). (B) For events occurring on chromosome XI in samples that retained trisomy for this chromosome throughout the experiment, the expected mutant frequency is 1/3, and again the observed mutant frequency did not differ from this expectation (t = –1.38, P = 0.17), but differed from 0.5 (t = –16.4, $P < 10^{-15}$). (C) For events occurring on chromosomes that became trisomic at some point during the experiment, there are three possible event types. Type I: mutation followed by trisomy of the non-mutant chromosome, giving allele frequency 1/3 with probability 1/4. Type II: mutation followed by trisomy of the mutant chromosome, giving allele frequency 2/3 with probability 1/4. Type III: trisomy followed by mutation, giving allele frequency 1/3 with probability 1/2. The expected overall allele frequency is therefore 0.417, and the observed mean frequency did not differ from this expectation (t = 0.53, P = 0.60), but differed from 0.5 (t = –2.29, $P < 0.05$). We expect

approximately 3/4 of events to have allele frequency <0.5, and the observed data did not differ from this expectation (18/25 events, binomial test, $P = 0.82$). Events could also occur at 2/3 frequency when there is trisomy followed by mutation followed by gene conversion, but such cases should be very rare (<0.1%) given the observed rates of homozygous SNMs in the experiment. One SNM occurred on a chromosome likely to be tetrasomic, with allele frequency 0.47 (not shown).

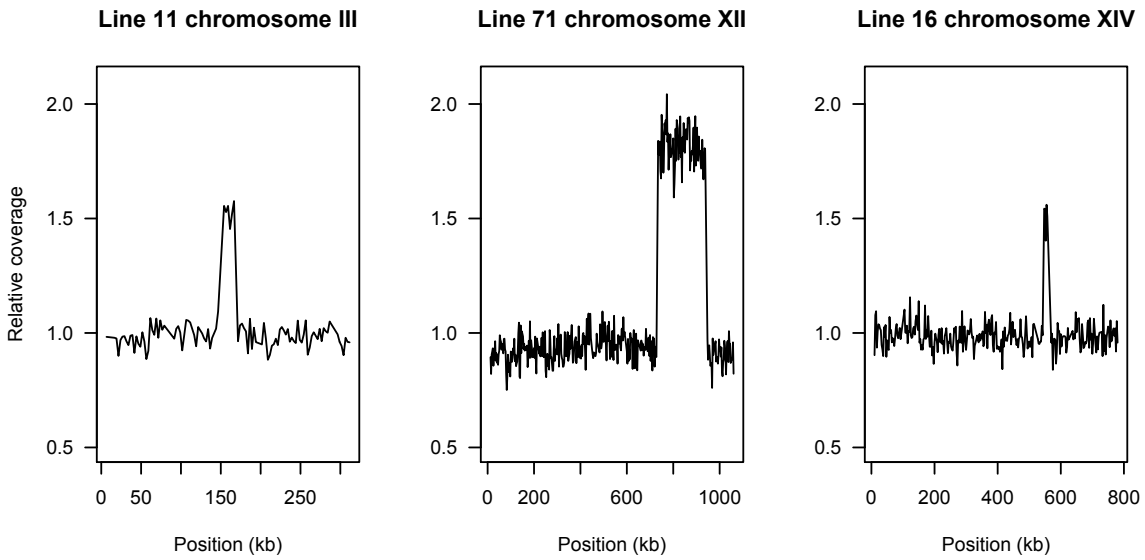

Figure S4. Large segmental duplications. Relative coverage in 2.5 kb windows for chromosomes and samples with suspected duplications. Here relative coverage represents the coverage ratio for a given window (mean coverage for sites in that window divided by median coverage for that sample and chromosome) relative to the median coverage ratio among all samples. In diploids a segmental duplication would be expected to result in relative coverage of 1.5. The case in line 11 appears to represent further duplication of a region that was already duplicated, relative to the reference genome, in the ancestor of the MA lines (note that this pattern is not visible in the relativized coverage profile shown). The samples plotted are all diploid; Line 11 is *RDH54*+ and the others are *rdh54*Δ. Approximate duplication lengths are 16.5 kb, 211.5 kb and 21.0 kb respectively.

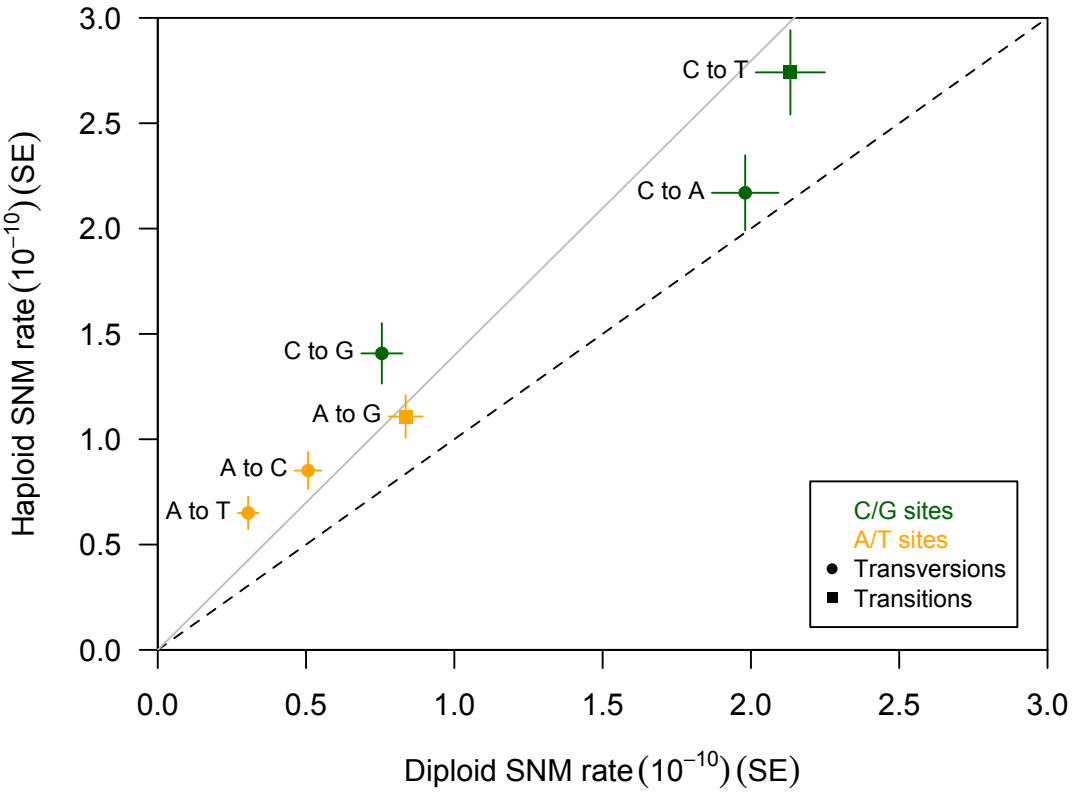

Figure S5. Haploid and diploid SNM rates per base pair per generation for each of the six mutation types, including the complementary change. The dashed line shows a 1:1 relationship; the solid gray line shows the slope corresponding to the overall haploid:diploid SNM rate ratio. Haploids had a higher SNM rate than diploids for all mutation types, but the relative rate increase was greater for some types of mutation (e.g., A to T) than others (e.g., C to A).

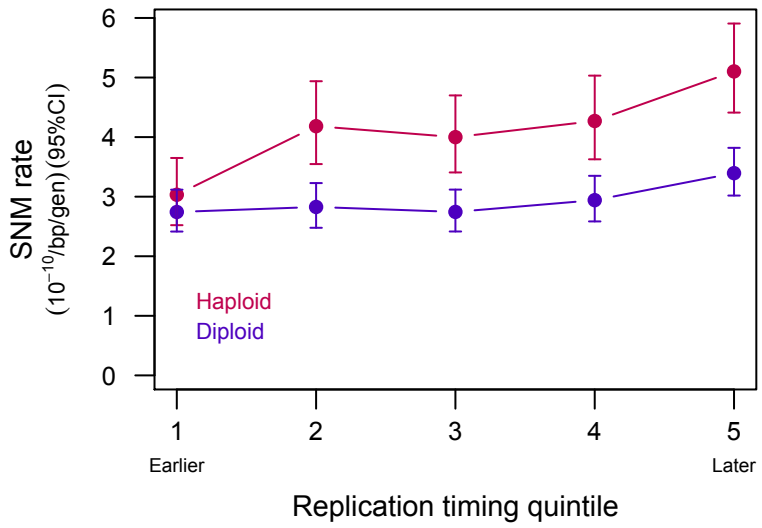

Figure S6. SNM rates in relation to replication time, for haploids and diploids. The genome was divided into five roughly-equal bins based on replication timing, and SNM rate calculated for sites in each bin. Haploids and diploids have a similar SNM rate in early-replicating sites (quintile 1), and haploids have a higher SNM rate in all other sites.

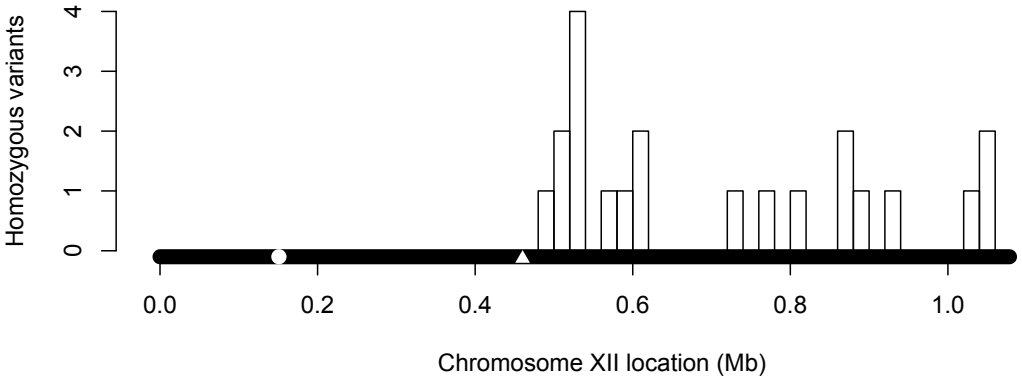

Figure S7. Locations of homozygous variants on chromosome XII relative to the centromere (white circle) and the region of rDNA repeats (white triangle). All homozygous variants on this chromosome were distal to the rDNA repeat region, and all occurred in unique MA lines. See Dataset S2 for homozygous variant locations.

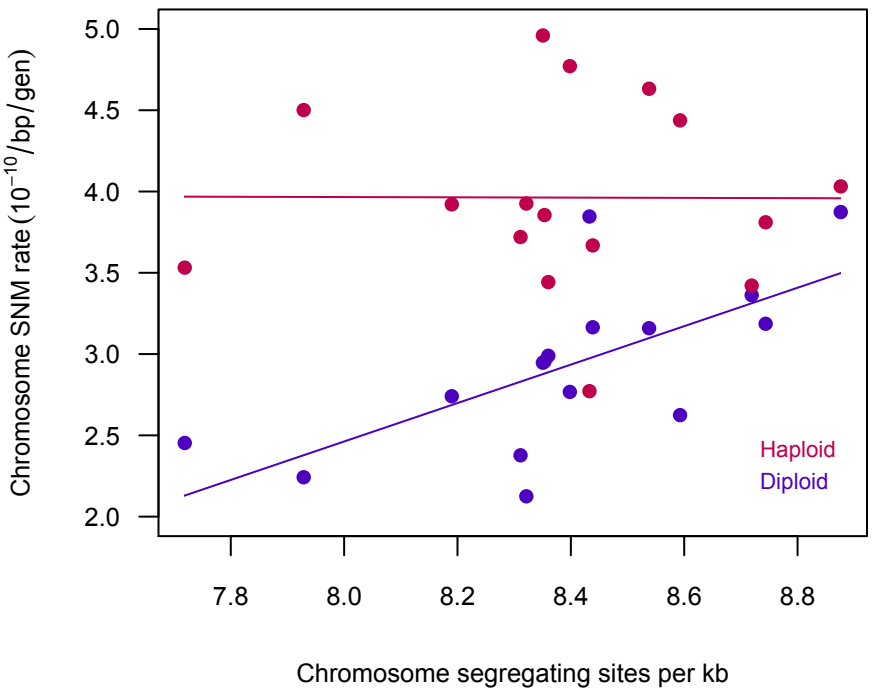

Figure S8. SNM rate on each chromosome in haploids or diploids versus number of polymorphic sites on that chromosome estimated from population samples. Polymorphism was correlated with SNM rate in diploids ($\rho = 0.79$, $P < 0.001$), but not haploids ($\rho = -0.05$, $P = 0.87$).

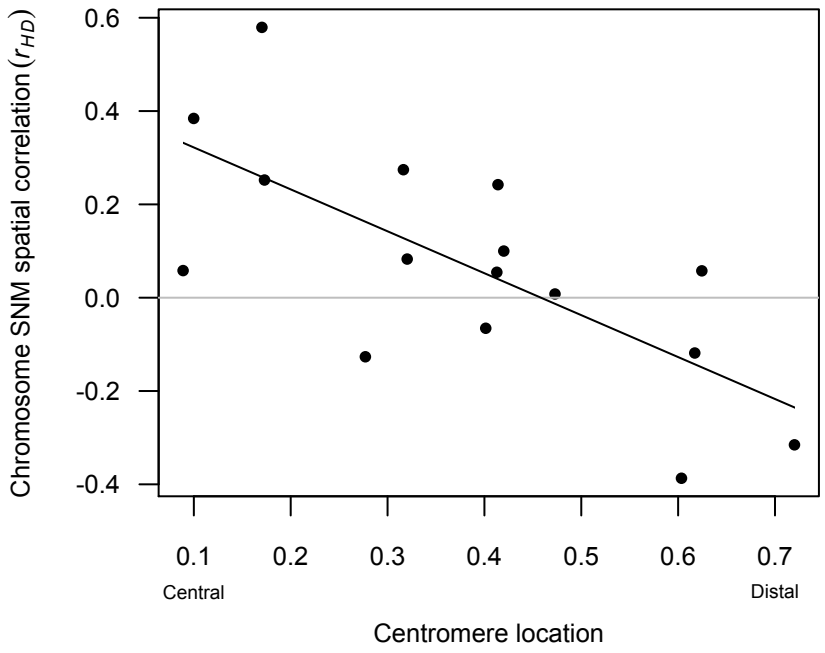

Figure S9. Spatial correlation of haploid versus diploid SNMs, r_{HD}, on each chromosome, plotted against centromere location relative to the chromosome center, $(L_q - L_p)/ (L_q + L_p)$. We calculated r_{HD} by estimating Gaussian kernel density of SNMs at 512 equally-spaced points along the chromosome with a smoothing bandwidth of 25 kb using the *R* function *density*, and then taking the Spearman correlation between haploid and diploid densities (alternative choices of bandwidth and correlation method gave similar results). Simulations (not shown) confirm that the expected r_{HD} for random sites is ~0, and that the observed correlation between r_{HD} and centromere location ($r = -0.70$, $P < 0.01$) is unlikely to occur if mutations are randomly distributed (5000 simulated datasets, $P < 0.01$).

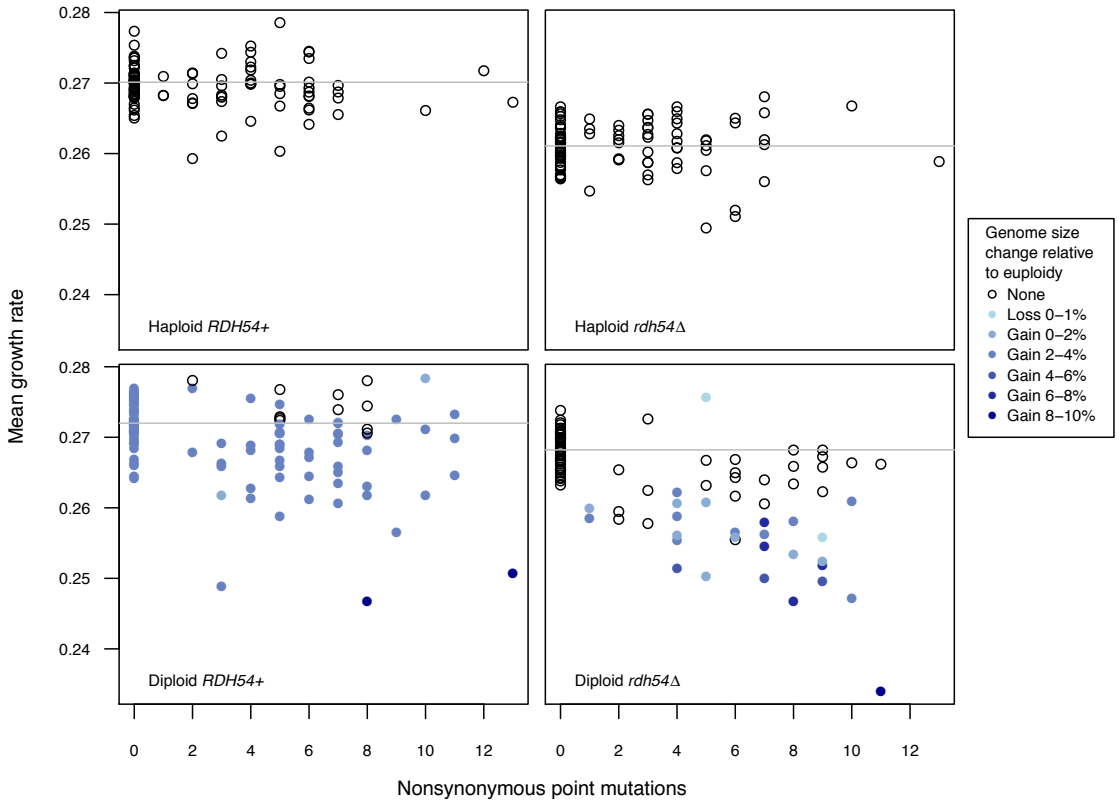

Figure S10. Mean growth rate (maximum rate of OD increase per hour) plotted against number of non-synonymous point mutations for each line. Colour codes indicate genome size change relative to euploidy. Ancestral control lines are plotted and assigned a value of zero for non-synonymous point mutations (only one MA line, from the haploid *RDH54*+ treatment, was found to have zero non-synonymous point mutations, and has a growth rate of approximately 0.27). The control mean is indicated with a horizontal gray line. Note that the diploid RDH54+ controls are trisomic for chromosome XI. Mean growth rate for a line is the expected value accounting for effects of initial OD, day and plate.

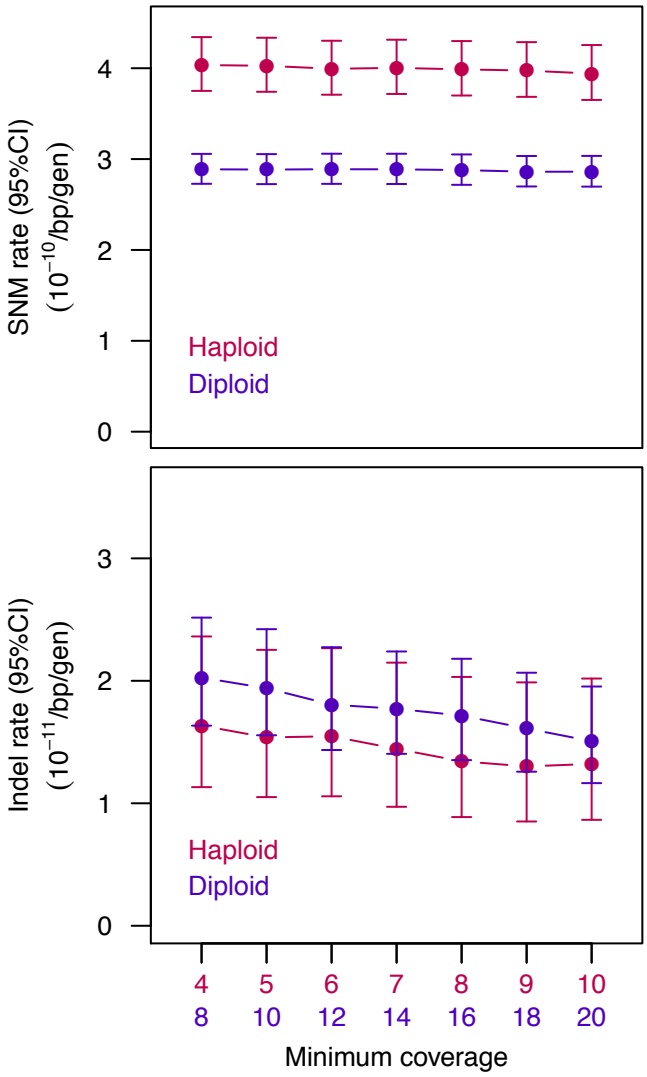

Figure S11. Mutation rate estimates using different minimum coverage cutoffs for SNMs (top) and indels (bottom). For each minimum coverage value, the number of mutations and callable sites was scored in haploids and diploids, where minimum diploid coverage equals two times minimum haploid coverage. Our primary analysis includes only sites in a given sample with coverage of at least 4 in haploids and at least 8 in diploids (leftmost points). Our findings of a significantly higher SNM rate in haploids than diploids and a non-significantly higher indel rate in diploids than haploids are robust to the minimum coverage cutoff used.

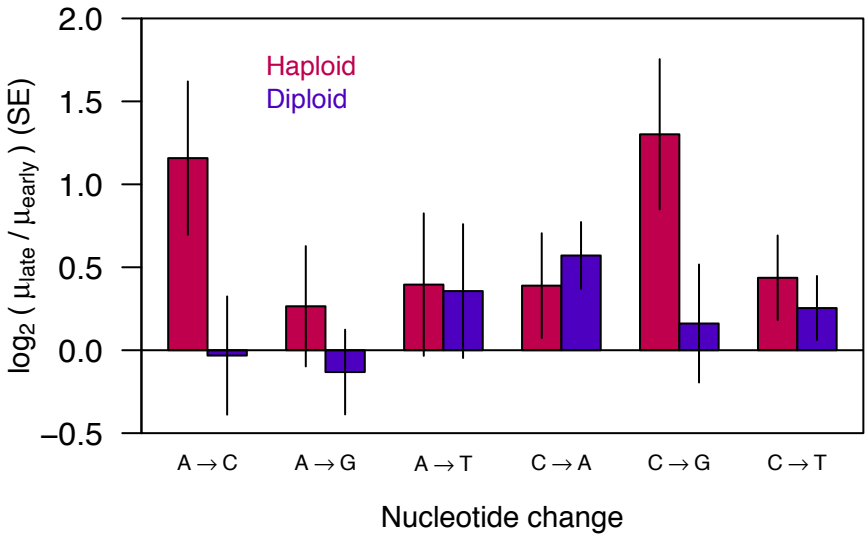

Figure S12. Changes in the rate of each type of SNM with replication timing. Bars indicate the log ratio of substitution rate for sites in the latest quintile of replication timing (μ_{late}) relative to the rate for sites in the earliest quintile (μ_{early}), accounting for detection power by site type and replication timing. Standard errors are from 5000 simulated datasets created with multinomial resampling of SNM types within groups. The difference in SNM spectrum between ploidy levels is significant in late-replicating regions (G = 13.1, df = 5, P < 0.05) but not early-replicating regions (G = 3.5, df = 5, P = 0.62), and this difference in G-statistics is unlikely to occur in the absence of an interaction between ploidy and replication timing (5000 datasets simulated assuming independence, P < 0.05). The increase in mutation rate in late-replicating regions is driven by significant differences (accounting for multiple tests) in A-to-C and C-to-G transversions in haploids, as well as C-to-A transversions in diploids (all P < 0.01).