The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast

Nathaniel P. Sharp[a,1], Linnea Sandell[a], Christopher G. James[a], and Sarah P. Otto[a]

[a]Department of Zoology, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Edited by Michael Lynch, Arizona State University, Tempe, AZ, and approved April 27, 2018 (received for review January 19, 2018)

By altering the dynamics of DNA replication and repair, alternative ploidy states may experience different rates and types of new mutations, leading to divergent evolutionary outcomes. We report a direct comparison of the genome-wide spectrum of spontaneous mutations arising in haploids and diploids following a mutation-accumulation experiment in the budding yeast *Saccharomyces cerevisiae*. Characterizing the number, types, locations, and effects of thousands of mutations revealed that haploids were more prone to single-nucleotide mutations (SNMs) and mitochondrial mutations, while larger structural changes were more common in diploids. Mutations were more likely to be detrimental in diploids, even after accounting for the large impact of structural changes, contrary to the prediction that mutations would have weaker effects, due to masking, in diploids. Haploidy is expected to reduce the opportunity for conservative DNA repair involving homologous chromosomes, increasing the insertion-deletion rate, but we found little support for this idea. Instead, haploids were more susceptible to SNMs in late-replicating genomic regions, resulting in a ploidy difference in the spectrum of substitutions. In diploids, we detect mutation rate variation among chromosomes in association with centromere location, a finding that is supported by published polymorphism data. Diploids are not simply doubled haploids; instead, our results predict that the spectrum of spontaneous mutations will substantially shape the dynamics of genome evolution in haploid and diploid populations.

aneuploidy | mitochondria | replication time | RDH54

Mutations play a critical role in evolution and adaptation, but these spontaneous genetic changes are often hazardous, reducing the health of individuals and imposing a genetic load on populations. While individual mutations are chance events, the biological processes that produce or prevent mutation have the potential to vary among genetic contexts, leading to consistent biases in the numbers, locations, and types of genetic changes that occur, affecting how genomes and populations evolve. One dimension of genomic variation that has broad potential to affect the mutation process is ploidy state, the number of homologous chromosome sets per cell. Ploidy varies among cells as a consequence of meiosis and syngamy and can also vary among individuals (e.g., in haplodiploid organisms). Transitions in predominant ploidy level have occurred many times across the tree of life (1, 2), but the evolutionary and ecological drivers and consequences of these changes are still unclear.

All else being equal, the genome-wide mutation rate of diploid cells should be twice that of haploid cells, as they have twice as many nucleotide sites with the potential to mutate. However, ploidy level is likely to have additional effects on the rate and spectrum of mutations. The presence of a homologous chromosome template allows DNA double-strand breaks (DSBs) to be repaired using conservative homology-directed pathways rather than competing error-prone end-joining pathways that generate insertion-deletion (indel) mutations (3). The continual access to homologous template DNA in diploid cells may therefore reduce the rate of indels relative to haploid cells. While diploidy may reduce susceptibility to certain forms of DNA damage, the presence of multiple genome copies may also increase the likelihood of spontaneous structural changes and rearrangements through nonhomologous crossing over (3). There are therefore reasons to expect differences in the mutational properties of alternative ploidy states.

Estimates of genome-wide mutation rates are available from mutation accumulation (MA) experiments with the budding yeast *Saccharomyces cerevisiae*, which can grow mitotically in either a haploid or diploid state. In these experiments, replicate lines are periodically subjected to single-cell bottlenecks, allowing mutations with mild and moderate effects to accumulate as though they were selectively neutral. Previous studies have examined either haploids or diploids, preventing direct comparison.

Genome-wide patterns of spontaneous mutation have been somewhat better characterized in diploid *S. cerevisiae* (4, 5) than in haploids, where sampling has been limited (6, 7). At face value, previous estimates suggest that the rate of single-nucleotide mutations (SNMs) per base pair may be greater in haploids than in diploids, and it has been proposed that yeast genomes are more stable in the diploid state (4, 5). However, these different inferred rates could reflect variation among laboratory strains, methodology, or environments. Additionally, in a previous study of haploids, all four lines examined by genome sequencing apparently became diploid spontaneously during the experiment (6), complicating inference of the mutation rate in haploid cells.

To formally investigate the impact of ploidy on the mutation process, we accumulated mutations in isogenic replicate haploid and diploid lines of *S. cerevisiae* derived from a common haploid ancestor. We maintained 220 lines under relaxed selection by

Significance

Organisms vary in the number of genome copies per cell: ploidy. By altering how DNA is replicated and repaired, ploidy may determine the number and types of mutations that arise, affecting how evolution proceeds. We sequenced the genomes of >200 replicate lines of yeast (*Saccharomyces cerevisiae*) with one versus two genome copies (haploid versus diploid) after accumulation of thousands of new mutations. Haploids were more susceptible to single-nucleotide mutations, particularly for DNA replicated later in the cell cycle, whereas large changes to genome structure were more common in diploids. Haploid and diploid populations will therefore have access to distinct kinds of genetic variation, contributing to differences in their evolutionary potential.

Author contributions: N.P.S., L.S., and S.P.O. designed research; N.P.S., L.S., and C.G.J. performed research; N.P.S. analyzed data; and N.P.S. and S.P.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the National Center for Biotechnology Information Sequence Read Archive, https://www.ncbi.nlm.nih.gov/sra/SRP139886 (accession no. SRP139886).

[1]To whom correspondence should be addressed. Email: sharp@zoology.ubc.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1801040115/-/DCSupplemental.

Published online May 14, 2018.

Table 1. Sample sizes

Metric	Haploid		Diploid	
	RDH54$^+$	*rdh54*Δ	*RDH54*$^+$	*rdh54*Δ
MA lines	53	53	63	51
Cell divisions per transfer	15.7	15.2	15.8	15.7
Transfers per line*	94.9	97.5	100.0	98.9
Median coverage per site per line (range†)	23 (16, 30)	22 (10, 35)	44 (20, 65)	44 (30, 62)
Callable nuclear sites per line‡	11,296,337	11,255,081	23,088,492	22,684,822
Callable mt sites per line	72,036	70,426	74,868	75,028

*Some lines were reinitialized partway through the procedure, and therefore underwent <100 transfers as unique lines.
†Minimum and maximum of median coverage among lines.
‡Incorporates ploidy, including ancestral trisomy of chromosome XI in some lines.

enforcing single-cell bottlenecks every day (~16 generations) for 100 d, for a total of ~336,000 generations across the experiment. To address the possible role of homology-directed DSB repair, we deleted the gene *RDH54* in the ancestor of half of the lines. There is evidence that this gene is essential for mitotic recombinational repair between homologous chromosomes in diploids but not between sister chromatids in haploids or diploids (8). Genome sequencing of our MA lines revealed over 2,000 mutation events, with clear impacts of ploidy on the rate, spectrum, locations, and fitness consequences of spontaneous mutations but little evidence that homology-directed DSB repair is a major factor in these differences.

Results

Growth Rates and Ploidy. To account for potential differences among treatments in rates of cell division, we measured the number of cells per colony in the MA lines throughout the bottlenecking procedure and used these values to calculate mutation rates per cell division (Table 1). We found small but significant treatment effects on growth rate, with a more rapid decline in diploid lines than in haploid lines over the course of the experiment (Fig. 1 *A* and *B*).

After 100 bottlenecks, we measured the growth rates of MA lines and ancestral controls in liquid culture, revealing that diploid MA lines, but not haploid MA lines, show reduced growth relative to ancestral controls (Fig. 1*C*), consistent with the measures of colony growth on plates during MA. All four groups of MA lines (ploidy × RDH54 status) show significant genetic variance for growth rate (all $\chi^2 > 24.8$, all $P < 10^{-6}$), whereas there was no significant genetic variance detected in any group of ancestral control lines (all $\chi^2 < 2.6$, all $P > 0.11$). The finding that diploid but not haploid lines show reduced growth following MA is surprising, as the effects of recessive deleterious mutations would be masked in diploids. We discuss this result in relation to the genomic data in *Growth Rates in Relation to Mutations*.

Haploid yeast are known to become diploid spontaneously (e.g., through defects in mitosis), and diploids can similarly change ploidy level. Such changes are commonly observed in yeast evolution experiments with large populations, but the spontaneous rate of ploidy change per cell division is unclear. Using flow cytometry, we found that all MA lines retained their original ploidy level at the end of the experiment [Dataset S1; confidence interval (CI) for rate of change from haploid to diploid: 0–2.3 × 10^{-5} per cell division; results on aneuploidy are discussed below]. This is in contrast to a previous finding (6), where four of four haploid MA lines became diploid over the course of ~4,800 generations [rate (95% CI): 20.8 (5.7, 53.3) × 10^{-5}]. This difference suggests either that the higher frequency of single-cell bottlenecks in our experiment (daily rather than every 3–4 d) reduced the opportunity for selection favoring diploidy or that yeast strains vary in their propensity to spontaneously change ploidy.

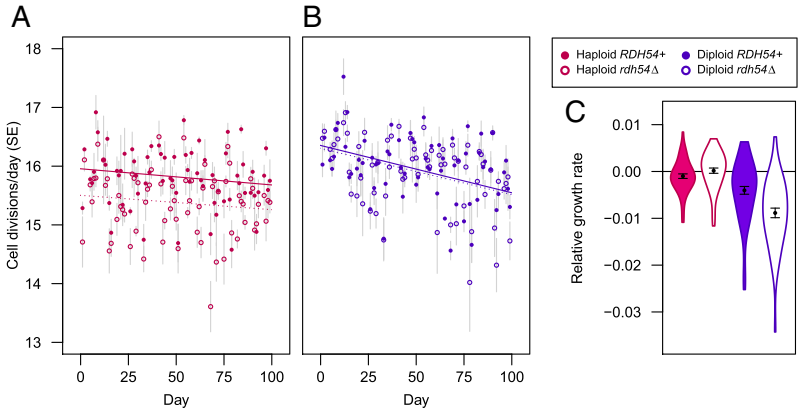

Fig. 1. Cell divisions during MA and growth rates following MA. Cell divisions per day (*A*, haploids; *B*, diploids) depended on the interaction of ploidy level, *RDH54* status, and time (>3,200 colonies measured; linear model, ploidy × *RDH54*: $P < 10^{-8}$; ploidy × time: $P < 0.01$). Solid (dashed) lines show linear regression for *RDH54*$^+$ (*rdh54*Δ) treatments. (*C*) Violin plots of maximum growth rates after 100 transfers relative to ancestral controls, accounting for block effects. Points (error bars) show means (SEs). Steeper declines in cell divisions per day in diploids relative to haploids during MA (*A* and *B*) are consistent with relative growth rate estimates following MA (linear mixed model based on 4,000 growth curves: MA × ploidy × *RDH54*: $P < 0.05$). Diploid MA lines have significantly reduced growth rates relative to the ancestor (MA: $P < 10^{-6}$, MA × *RDH54*: $P < 0.05$), unlike haploid lines (MA: $P = 0.64$). Growth-rate data for individual lines are provided in Dataset S2.

We observed changes in the mating behavior of three MA lines and describe the likely genetic basis for these changes in *SI Appendix*.

Point Mutations. We detected >2,000 point mutations (Dataset S2). Mutation rates, accounting for numbers of cell divisions and callable sites, are shown in Fig. 2. There was no effect of mating type (MAT) on mutation rates within haploids (*SI Appendix*, Fig. S1), and so we pooled the data for haploid lines of the two mating types (MATa and MATα) throughout our analyses. We detected a significant effect of ploidy level on the rate of SNMs (binomial test: $P < 10^{-11}$), with no effect of *RDH54* status ($P > 0.24$ in either ploidy level). The per-nucleotide SNM rate was 40% higher in haploids (4.04×10^{-10}, 95% CI: 3.75–4.34×10^{-10}) than in diploids (2.89×10^{-10}, 95% CI: 2.73–3.06×10^{-10}). Thus, although diploids have twice as many nucleotide sites as haploids, they incur only 1.43-fold as many SNMs per cell division.

Because diploidy provides greater access to homologous DSB repair, we expected to see a higher rate of indel mutations in haploids and in diploids lacking the *RDH54* gene. However, the indel rate did not depend on ploidy level (binomial test: $P = 0.52$) or *RDH54* status ($P > 0.58$ in either ploidy level; Fig. 2). In fact, our point estimate of the per-nucleotide rate of indels is 24% greater in diploids (2.03×10^{-11}) than in haploids (1.63×10^{-11}), such that the indel rate per cell division is 2.48-fold higher in diploids. There was a nonsignificant tendency toward deletions among indel events (62 deletions and 50 insertions; binomial test: $P = 0.30$).

We considered whether selection during the experiment might have affected rates of MA. Accounting for power to detect mutations in genic and nongenic regions and assuming that genic and nongenic regions are equally likely to mutate, there is no evidence that SNMs were less likely to occur in genes than expected (observed rate: 74.0%, expected rate: 73.9%; binomial test: $P = 0.92$) or less likely to be nonsynonymous than expected (observed rate: 76.7%, expected rate: 76.1%; binomial test: $P = 0.62$), with no difference in these rates between ploidy or *RDH54* levels. There is therefore no evidence that the accumulation of SNMs was influenced by selection in our experiment.

By contrast, indels were found in genes less often than expected (observed rate: 46.4%, expected rate: 73.9%; binomial test: $P < 10^{-9}$), as observed in other datasets of this kind (5). Selection acting more effectively against indels in haploids could mask a true difference in indel rate. However, the fraction of indels that accumulated in genes does not differ significantly between ploidy levels (Fisher's exact test: odds ratio = 1.10, $P = 0.83$), and there is no indication that nongenic indel rates differed by RDH54 status or ploidy (*SI Appendix*, Fig. S2). It is therefore not apparent that the pattern of indel accumulation among treatment groups in our experiment was driven by selection against genic indels. The deficit of indels in genes could also stem from differences in sequence complexity between genic and nongenic regions, rather than selection. We observed indels >20-fold more frequently in regions identified by RepeatMasker (9) as simple repeats or a low-complexity sequence (binomial test: $P < 10^{-15}$), and such regions were >2.8-fold more common outside of genes (odds ratio = 0.34, $P < 10^{-15}$). Excluding these regions, we still find that indels are less common in genes than expected (binomial test: $P < 10^{-8}$), but this simple approach is unlikely to account for all repetitive sequences.

As in previous studies where MA lines have been sequenced, we find evidence for some substitutions occurring in close proximity to one another, which we categorize as multinucleotide mutations (MNMs) (10). As with SNMs, the MNM rate was significantly higher in haploids (0.73×10^{-11}) than diploids (0.29×10^{-11}; Fig. 2; binomial test: $P < 0.05$), suggesting that replication errors associated with ploidy generate both SNMs and MNMs. This difference is particularly pronounced in the *rdh54*Δ haploids (Fig. 2; binomial test of RDH54 effect within haploids: $P = 0.056$), but the ploidy × RDH54 interaction is nonsignificant (power-adjusted goodness-of-fit test: $G = 2.0$, $P = 0.16$).

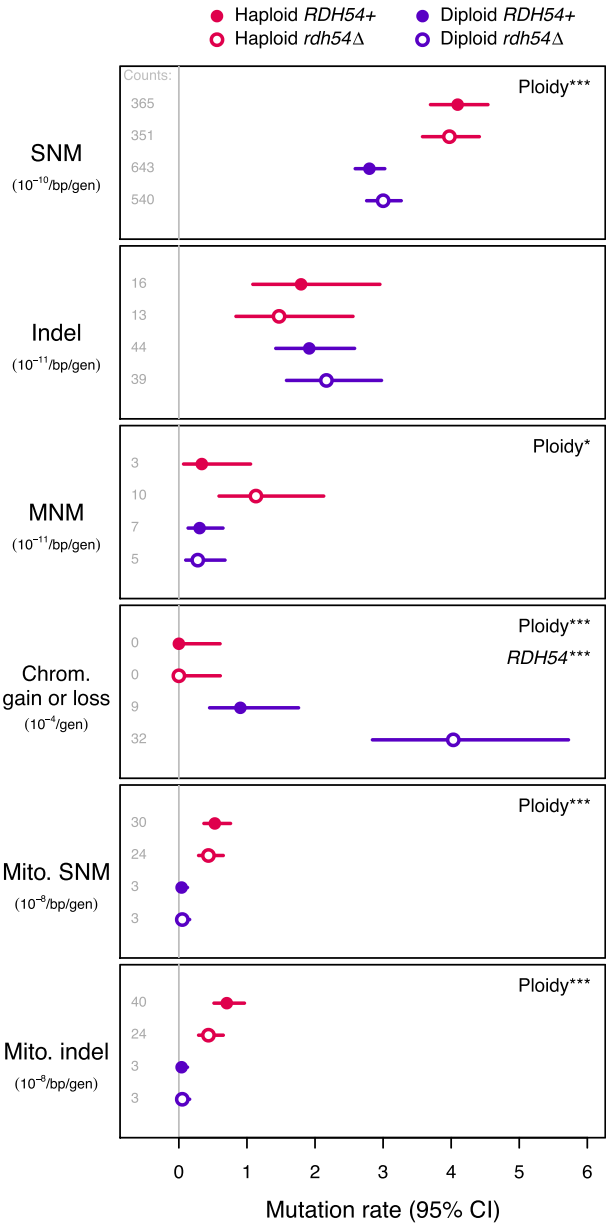

Fig. 2. Mutation rates in each group of MA lines, with 95% CIs. Panels show SNMs, indels, MNMs, chromosome (Chrom.) gains and losses, mt (Mito.) SNMs, and Mito. indels. Rates represent events per base pair (bp) per generation (gen), except for whole-chromosome gains and losses (events per generation). For mt events, we consider all treatments effectively haploid. The absolute numbers of events observed ("Counts") are given at the left of each panel; note that detection power differs among groups such that rates are not necessarily proportional to mutation counts. Statistically significant treatment effects are noted at the right of each panel (binomial tests accounting for detection power: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). A numeric summary of rates and CIs is provided in Dataset S2.

While asexual diploids are expected to adapt more slowly because beneficial mutations are partially masked, the strength of this effect depends on the rate at which mutations become

homozygous (11), due to mitotic crossing over, noncrossover gene conversion, or chromosome loss and duplication (3, 12). We identified 38 homozygous SNMs and one homozygous indel in diploid lines, reflecting loss of heterozygosity (LOH) events. The rate of homozygous mutations did not differ between *RDH54* treatments (binomial test: $P = 0.63$). For six homozygous mutations, another mutation occurred distally on the same chromosome arm in the same MA line, and of these, three were also homozygous, which is a higher rate than expected if all LOH events occurred independently (binomial test: $P < 0.001$). These pairs of homozygous mutations were separated by 132–451 kb, suggesting that a substantial fraction of LOH events are due to mitotic crossing over or chromosome loss/duplication rather than relatively short gene conversion tracts.

Large-Scale Mutations. We used sequencing coverage across the genome to detect changes in chromosome copy number (aneuploidy), which result from mitotic nondisjunction. Forty-nine of 63 *RDH54*$^+$ diploid lines were found to carry three copies of chromosome XI, which we established to be an ancestral polymorphism in that group of lines (*Materials and Methods*). We also observed 41 de novo chromosome gains and losses, all within the diploid MA lines (Fig. 2). Chromosome gains (35 cases of trisomy and two cases of tetrasomy) were found more often than chromosome losses (four losses, including a loss of a trisomic chromosome XI restoring euploidy; binomial test: $P < 10^{-6}$; Fig. 3*A*), as observed in a previous MA experiment (5). While the rarity of chromosome losses relative to gains may reflect stronger selection against monosomy in diploids or a higher chance of reversion to euploidy from monosomy in MA experiments, our observed rates of chromosome loss are generally consistent with previous estimates obtained using other methods (reviewed in ref. 3). The derived versus ancestral allele depths for substitutions conformed well to the expected patterns on aneuploid and nonaneuploid chromosomes (*SI Appendix*, Fig. S3), confirming that the chromosome XI trisomy occurred before MA and that other aneuploidies occurred uniformly throughout MA.

While the rate of aneuploidy in *RDH54*$^+$ diploids (0.90×10^{-4}) is very similar to previous observations from diploid yeast (1.04×10^{-4}) (5), we found that *rdh54*Δ elevated the rate of aneuploidy fourfold overall and 4.5-fold within diploids (Fig. 2; binomial test: $P < 10^{-4}$), indicating that *RDH54* plays a role in mitotic chromosome segregation in addition to its role in meiosis (8, 13), as suggested previously (14). Given the absence of aneuploidy in haploids, we cannot determine whether the increase in aneuploidy in *rdh54*Δ lines also applies to haploids or is unique to diploids.

Accounting for the range of haploid aneuploidy rates that is statistically consistent with finding no events and comparing this with the inferred rate in diploids, we find that the rate of aneuploidy is significantly higher in diploids on a per-cell-division basis (Fig. 2; binomial test: $P < 10^{-11}$). This pattern persists when excluding chromosome losses, which would be lethal in haploids (binomial test: $P < 10^{-10}$) and when excluding *rdh54*Δ lines (binomial test: $P < 0.05$). Accounting for the number of chromosomes per cell that could potentially increase in copy number (twofold higher in diploids), the effect of ploidy on the rate of chromosome gains remains significant overall (binomial test: $P < 10^{-5}$). Considering only gains in the *RDH54*$^+$ lines, however, we cannot reject the hypothesis that the opportunity for nondisjunction is simply twofold higher in diploids than in haploids (binomial test: $P = 0.20$). In any case, our data indicate that diploids experience at least twice the rate of whole-chromosome changes per cell division compared with haploids, in contrast to their lower rate of point mutation (discussed above).

The probability of aneuploidy was not uniform across chromosomes (Fig. 3*A*), and there was a negative but nonsignificant correlation between the rate of gain or loss and chromosome length (all lines: $r = -0.41$, $P = 0.11$; *RDH54*$^+$ lines: $r = -0.36$, $P = 0.17$; *rdh54*Δ lines: $r = -0.41$, $P = 0.11$), supporting the view that chromosome size is not the sole determinant of aneuploidy rates (3, 5). Variation in aneuploidy rates among chromosomes in our experiment was correlated with aneuploidy rates observed previously in diploids exhibiting chromosome instability ($r = 0.57$, $P < 0.05$) (15). We found that growth rates were highly correlated with genome size among aneuploid lines (Fig. 3*B*).

In addition to aneuploidy, coverage profiles revealed three cases of large segmental duplications, all heterozygous in diploids

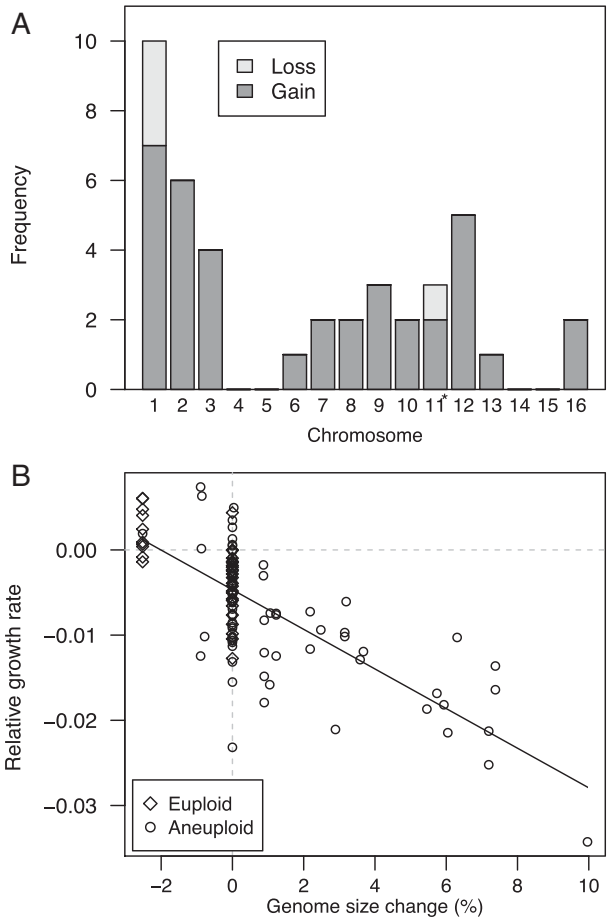

Fig. 3. Gains and losses of each chromosome during MA and the fitness consequences of aneuploidy. (*A*) There was no evidence that the distribution of aneuploidy events among chromosomes depended on *RDH54* type (Fisher's exact test: $P = 0.56$), but with few events in *RDH54*$^+$ lines, we have low power to test this hypothesis. Pooling data between *RDH54* types, gains were more common than losses (binomial test: $P < 10^{-6}$) and rates vary among chromosomes ($\chi^2 = 42.1$, simulated $P < 0.001$). Several cases of trisomy for chromosome 11 in *RDH54*$^+$ diploids were determined to be ancestral and are not scored as gains; in one case, the extra copy was lost, restoring euploidy, and this is scored here as a loss. Aneuploid line identifications are provided in Dataset S2. (*B*) Genome size relative to controls was negatively correlated with maximum growth rate relative to controls in diploid MA lines ($r = -0.74$, df = 112, $P < 10^{-15}$). Diploid *RDH54*$^+$ controls had trisomy for chromosome 11, so MA lines with only this trisomy were scored as having zero genome size change and MA lines without this trisomy were scored as having a genome size reduction unless other aneuploidy was present. This correlation persists when only aneuploid lines are considered ($r = -0.70$, df = 78, $P < 10^{-12}$) or when lines with no change are excluded ($r = -0.84$, df = 45, $P < 10^{-12}$).

EVOLUTION

(*SI Appendix*, Fig. S4), ranging in size from 17 kb to 211 kb. The approximate breakpoints of all three duplications are closer to known Ty elements (mean distance = 700 bp, range: 100–1,662 bp) than expected by chance (simulation of 10^4 sets of six breakpoints; mean distance to Ty ≤ 700; $P < 10^{-4}$), supporting the view that Ty elements often anchor large duplications in yeast (16–19). Our estimate of the large duplication rate in diploids is 1.68 (95% CI: 0.34, 5.24) × 10^{-5} per cell division, consistent with a previous observation of diploid MA lines (~1.00 × 10^{-5}) (5).

The presence of mutations could increase the subsequent mutation rate by directly altering DNA repair genes, by increasing genome instability (3, 7), or if genetic quality affects DNA repair (20). However, we found no evidence for an effect of aneuploidy on the point mutation rate and no evidence for additional variation in the number of mutations per line beyond that expected under a Poisson distribution (i.e., no evidence of overdispersion) (*SI Appendix*).

Mitochondrial Mutations. The cellular environment, replication mechanisms and nucleotide composition of mitochondrial (mt) genomes are distinct from those of nuclear genomes, but it is not clear how nuclear genome ploidy might interact with mt mutation. Unexpectedly, we found that the rate of mt mutation was an order of magnitude higher in haploids (SNM: 4.82 × 10^{-9}, indel: 5.71 × 10^{-9}) than in diploids (SNM: 4.47 × 10^{-10}, indel: 4.47 × 10^{-10}; Fig. 2 and *SI Appendix*, Table S1). This pattern remained when we relaxed our filtering criteria to detect putative heteroplasmic mutations, which are not fixed within an MA line (*SI Appendix*). Mt mutation rate estimates based on fewer events from a previous study of four haploid MA lines (which became diploid during MA) are even higher than the haploid rates we observed (SNM: 12.2 × 10^{-9}, indel: 10.4 × 10^{-9}) (6). As in this previous study, we found that the haploid substitution rate was higher in the mt genome than in the nuclear genome (11.9-fold difference, binomial test: $P < 10^{-15}$), but this was not the case for diploids (1.6-fold difference, binomial test: $P = 0.30$).

As with nuclear point mutations, there was no evidence that mt SNMs occurred in genes less often than expected (observed proportion: 0.27, expected: 0.35; binomial test: $P = 0.23$), but mt indels were less likely to accumulate in genes than expected (observed: 0.19; binomial test: $P < 0.01$). The fraction of events in genes did not differ between ploidy levels for either SNMs ($G = 0.32$, $P = 0.57$) or indels ($G = 0.07$, $P = 0.79$), and so there is no indication that the elevated rate of mt mutations in haploids relative to diploids is the result of more effective selection against such events in diploids. Small insertions and deletions were about equally prevalent among mt events (37 deletions, 33 insertions; binomial test: $P = 0.72$).

Mt deletions or other mutations can sometimes result in a "petite" phenotype due to defects in respiratory function. Although we attempted to prevent the accumulation of such variants, two haploid lines had a petite phenotype by the end of the experiment. An examination of sequencing coverage in these lines indicated large deletions in COX1 with very similar breakpoints (deleted locations: ~16,438–24,230 and 16,435–24,592). Mt deletions with these approximate breakpoints are known to occur frequently, possibly due to autocatalytic activity (21). Additionally, two haploid lines showed large mt deletions in nongenic regions based on coverage (deleted locations: ~4,824–5,751 and 80,559–83,041). An additional haploid line with a missense mutation in COX1 also showed poor respiration. Thus, in the mt genome, both point mutations and structural changes were observed more often in haploids, unlike in the nuclear genome, where structural changes were observed more often in diploids.

Spectrum of Nucleotide Changes. Haploids and diploids differed in the spectrum of SNMs (Fig. 4*A*). In particular, A-to-T and C-to-G changes were more common among haploid SNMs than among diploid SNMs, whereas the reverse was true for C-to-A changes. Overall, mutations at A/T sites made up a larger fraction of SNMs in haploids than in diploids (Fig. 4*B*), although all six types of substitution occurred more frequently in haploids than in diploids (*SI Appendix*, Fig. S5). Such mutational biases may contribute to different substitution spectra in adapting haploid and diploid populations (22).

The transition/transversion ratio did not differ significantly between ploidy levels (haploid = 0.75, diploid = 0.86; odds ratio = 0.88; $P = 0.18$), and the overall ratio [0.82 (95% CI: 0.75–0.90)] fell between previous estimates for initially haploid MA lines (0.62) (6) and diploid MA lines (0.95) (5).

SNM rates were affected by adjacent nucleotide context at C/G sites but not at A/T sites, with no differences in context effects between ploidy levels (Fig. 4*B*). In our experiment, both haploids and diploids show an elevated substitution rate at CpG sites compared with C/G sites in other contexts (binomial test: $P < 10^{-11}$; Fig. 4*B*), with an approximately twofold elevation in the rate of C-to-T mutations in this context (binomial test: $P < 10^{-14}$). Although *S. cerevisiae* shows little evidence for cytosine methylation (but see ref. 23), which can increase the frequency of C-to-T mutations in CpG contexts (24), such an increase has been observed in both diploid *S. cerevisiae* (5) and haploid *Schizosaccharomyces pombe* (25, 26).

Genomic Locations of Mutations. Previous studies have found evidence for increased mutation in late-replicating genomic regions in haploids (27) and nonsignificant trends in diploids (5). Here, we use published information on replication across the yeast genome (28) to compare the effect of replication timing between ploidy levels. Replication timing across the genome is very similar between ploidy levels (28), but we find that SNMs were more likely to occur in later-replicating regions in haploids, whereas there was no effect of replication timing on the rate of SNMs in diploids (Fig. 5*A*). This effect of replication timing in haploids at least partly explains the overall difference in SNM rates between ploidy levels: We found that haploid and diploid SNM rates were similar within early-replicating genomic regions but diverged in later-replicating regions (*SI Appendix*, Fig. S6).

We found that substitutions, but not indels, occurred further from the centromere than expected by chance (Fig. 5*B*), but neither ploidy nor *RDH54* status affected the average distance of mutations from the centromere (linear models: all $|t| < 0.99$, all $P > 0.32$). Homozygous SNMs were especially likely to occur far from the centromere (Fig. 5*B*), which is expected if LOH is caused by mitotic crossing over, although noncrossover gene conversion has also been found to be more frequent further from the centromere (20). Homozygous variants occurred nonrandomly among chromosomes ($\chi^2 = 114$, $P < 10^{-15}$), and the distribution differs from that of heterozygous variants ($\chi^2 = 143$, $P < 10^{-15}$). This pattern is driven almost entirely by an excess of homozygous variants on chromosome XII (21 of the 39 events, all in different lines; binomial test: $P < 10^{-12}$), all distal to the rDNA tandem repeat region (*SI Appendix*, Fig. S7), which is known to exhibit a high LOH rate (29, 30). Excluding events on chromosome XII, homozygous variants still occur further from the centromere than expected by chance (Wilcoxon and Kolmogorov–Smirnov tests: $P < 0.01$).

Previous studies have found evidence that mutation rates are correlated with the guanine-cytosine (GC) content of surrounding regions (20, 31). While we find no correlation between SNM rate and GC content (in accordance with ref. 5), we find that indels are associated with low GC content (in accordance with ref. 20), whereas MNMs are associated with high GC content (Fig. 5*C*). These associations diminish rapidly with distance from the focal site.

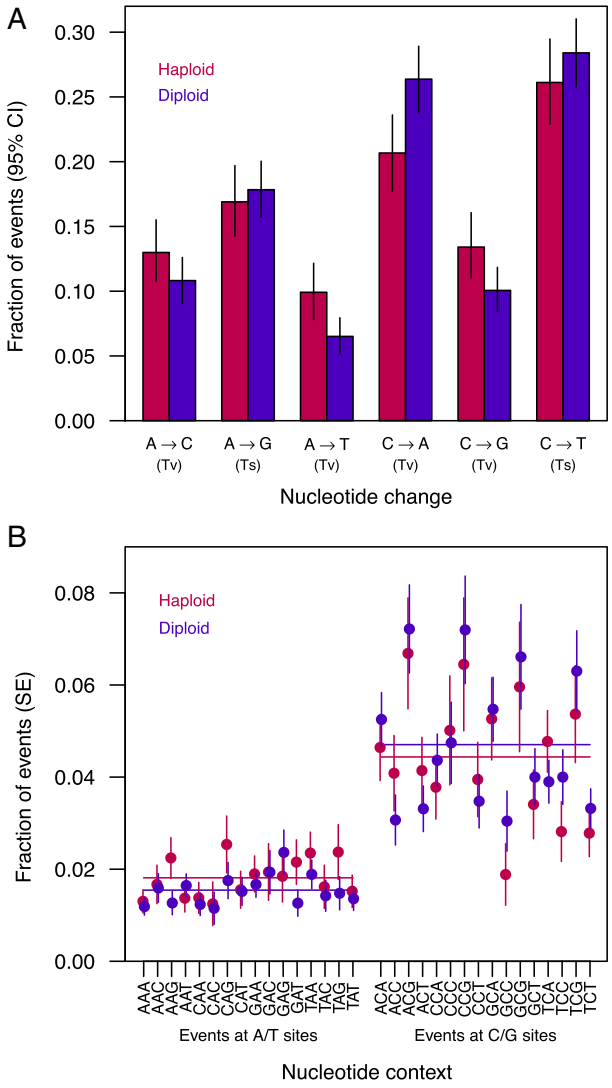

Fig. 4. Spectrum of nuclear SNMs. (*A*) Fraction of SNMs of each type, including complementary changes. Ts, transition; Tv, transversion. *RDH54* status did not significantly affect the SNM spectrum ($G = 4.5$, $P = 0.48$), but the SNM spectrum differed between ploidy levels ($G = 19.8$, $P < 0.01$), particularly among transversion mutations ($G = 17.9$, $P < 0.001$). By contrast, the spectrum of transition mutations and the overall transition-to-transversion ratio did not vary significantly by ploidy ($G = 0.04$, $P = 0.84$ and odds ratio = 0.88, $P = 0.18$, respectively). (*B*) Fraction of SNMs in each 3-bp context, including the complementary context, centered on the focal site, accounting for the frequency of each context in the genome. While the context of SNMs did not differ significantly between haploids and diploids when considering all contexts ($G = 21.5$, $P = 0.90$), the fraction of mutations at A/T versus C/G sites differed between ploidy levels (odds ratio = 1.22, $P < 0.05$; horizontal lines show means across A/T or C/G sites for each ploidy level). The overall rate of SNMs depended on the 3-bp context at G/C sites ($G = 81.9$, $P < 10^{-10}$) but not at A/T sites ($G = 20.9$, $P = 0.14$). The mutation rate in XCG contexts (where X is any base) was elevated relative to the rate for C/G sites in other contexts (binomial test: $P < 10^{-11}$).

While the number of SNMs on a given chromosome was highly correlated with callable chromosome length (haploids: $r = 0.96$, $P < 10^{-8}$; diploids: $r = 0.96$, $P < 10^{-8}$), there was some evidence for residual variation (Fig. 5*D*), and we found that diploid SNM rates were significantly lower on chromosomes where the centromere is relatively distal, whereas this correlation was absent, and possibly reversed, in haploids (Fig. 5*D*). We reasoned that if this unexpected result reflects genuine mutation rate variation among chromosomes, the pattern might also be evident in the locations of polymorphisms. We found that polymorphic sites in wild and domesticated yeast isolates (32) coincided with SNM sites more often than expected by chance (odds ratio = 1.52, $P < 0.05$) and that polymorphism rates by chromosome were more strongly correlated with our diploid SNM rates ($\rho = 0.79$, 95% CI: 0.45–0.95, $P < 0.001$; *SI Appendix*, Fig. S8) than our haploid SNM rates ($\rho = -0.05$, 95% CI: −0.50 to 0.41, $P = 0.87$), as expected if polymorphism reflects mutation rates and given that *S. cerevisiae* is generally diploid. As with SNMs in diploids, polymorphism rates tend to be negatively correlated with centromere location ($\rho = -0.47$, $P = 0.066$).

Examining the spatial pattern of SNMs within chromosomes, we found that the extent of spatial correlation between ploidy levels was itself negatively correlated with centromere location ($r = -0.70$, $P < 0.01$; *SI Appendix*, Fig. S9), which is unlikely to occur by chance (5,000 simulated datasets; $P < 0.01$). This suggests that the mutation rate in diploids is reduced in particular regions of chromosomes with distal centromeres. Additional possible explanatory factors are addressed in *SI Appendix*. The relative position of centromeres will affect the physical location of chromosome arms within the nucleus, as well as interactions between homologous loci (33–35), which might differentially affect haploid and diploid genome stability.

Growth Rates in Relation to Mutations. In our study, MA led to reduced growth in diploids but not haploids (Fig. 1*C*). This difference was driven, at least in part, by a particular class of mutations, aneuploidy events, that arose in diploid but not haploid lines (Fig. 3*B*). Accounting for these large-effect variants, including the presence of chromosome XI trisomy in the ancestral diploid *RDH54*$^{+}$ control, we modeled the effects of genome size and number of nonsynonymous nuclear and mt point mutations using mixed models (*SI Appendix*, Fig. S10, and details are provided in *SI Appendix*). We confirmed a highly significant effect of genome size change ($\chi^2 = 291.37$, $P < 10^{-15}$) and detected an interaction between ploidy and point mutations ($\chi^2 = 5.91$, $P < 0.05$). Separate models for haploids and diploids showed a negative but nonsignificant effect of point mutations in haploids ($\chi^2 = 0.32$, $P = 0.57$) and a significant negative effect in diploids ($\chi^2 = 19.56$, $P < 10^{-5}$).

These results, as well as other modeling approaches (*SI Appendix*), indicate that the accumulated point mutations had stronger negative effects on diploids than haploids, controlling for the effects of aneuploidy. This unexpected pattern could arise if the accumulation of deleterious alleles was opposed by selection during MA in haploids but not in diploids, although we did not find evidence of selection, at least among the SNMs, when examining the fraction of genic or nonsynonymous events. Furthermore, the effective population sizes of our lines should permit the accumulation of mutations with effects less than ~12%, which is more extreme than the largest fitness decline we observed (~3%), suggesting that the vast majority of mutations had weak effects and would accumulate as though they were neutral. Even if selection had no influence on which mutations accumulated, it is possible that the types or locations of point mutations that accumulated in diploids had more severe fitness effects on average. Alternatively, it is possible that beneficial mutations account for the difference in average fitness between haploid and diploid MA lines, as the initial strain was not given the opportunity to adapt to the specific conditions of the growth rate assay. The haploid MA lines exhibited significant genetic variance in growth rate but no significant change in average growth rate, consistent with the presence of beneficial mutations (36–38); beneficial mutations could occur at both ploidy levels but be partially recessive in diploids, which would account for the

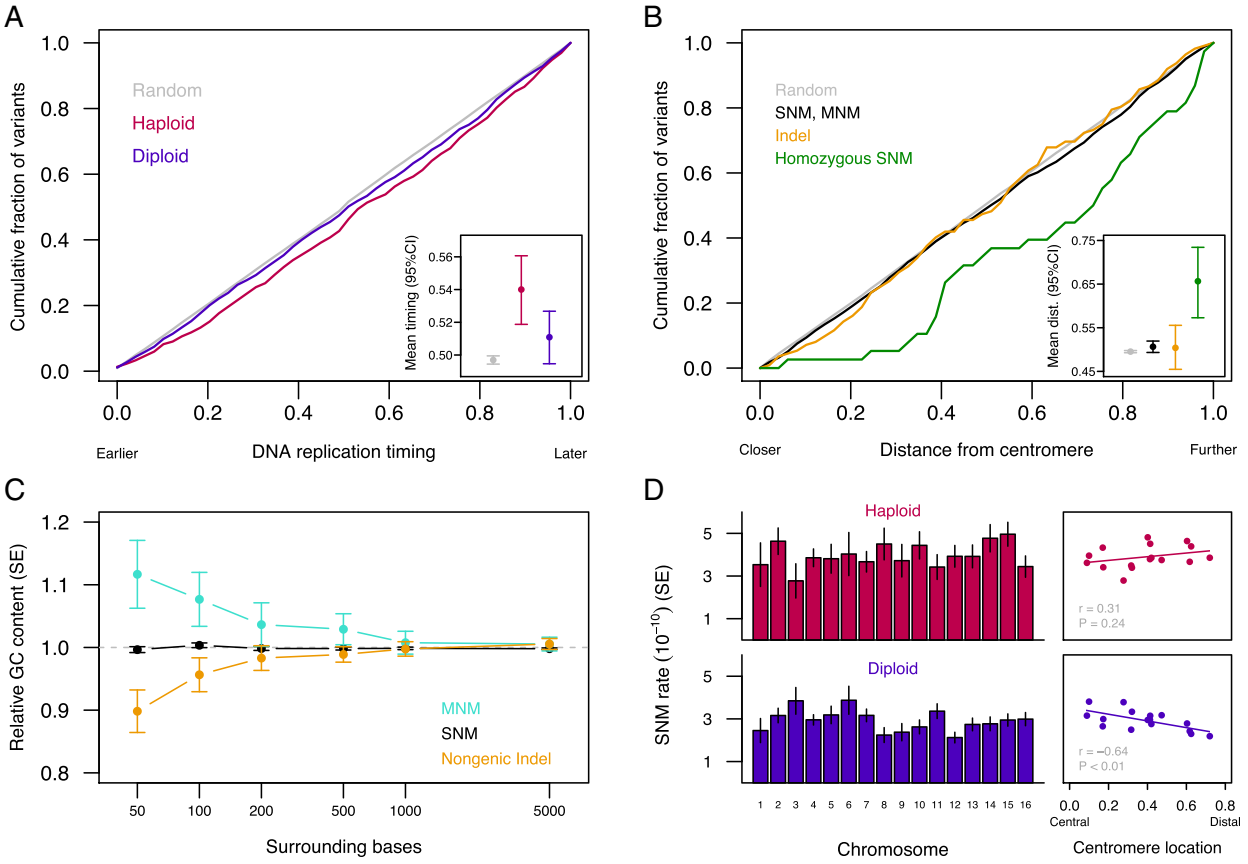

Fig. 5. Genomic context of nuclear mutations. (*A*) DNA replication timing (data from ref. 28) of mutated sites differed significantly between haploids (red) and diploids (blue) (Wilcoxon test: $P < 0.05$). Haploid mutations arose in later-replicating positions relative to random callable sites (Wilcoxon test: $P < 10^{-4}$), whereas diploid mutations did not differ significantly from the random expectation (Wilcoxon test: $P = 0.09$). (*Inset*) Mean replication timing for each group (also *SI Appendix*, Fig. S6). (*B*) Mutations versus distance from the centromere, as a fraction of maximum possible distance. All event types except indels occurred further from the centromere than random callable sites (Wilcoxon test, SNM and MNM: $P < 0.05$, Wilcoxon test, homozygous: $P < 0.001$), and homozygous diploid SNMs are further from the centromere than heterozygous SNMs (Wilcoxon test: $P < 0.001$). (*Inset*) Mean distance from the centromere for each event type. (*C*) GC content surrounding mutations relative to GC content surrounding equivalent random sites (A/T or C/G sites for substitutions, genic or nongenic sites for all mutation types). We consider only nongenic indels here because of possible selection against genic indels. MNM values represent the mean for the component sites. MNMs were associated with higher GC content within 50 bp (bootstrap $P < 0.05$), whereas nongenic indels were associated with lower GC content within 50 bp (bootstrap $P < 0.01$). (*D*) Variation in SNM rate among chromosomes (*Left*; diploids: $G = 23.7$, $P = 0.07$; haploids: $G = 12.0$, $P = 0.68$) was correlated with the relative distance between the centromere and the chromosome midpoint in diploids (*Right*), and this correlation differs significantly between ploidy levels (bootstrap $P < 0.01$). Centromere location was calculated as $(L_q - L_p)/(L_q + L_p)$, where L_q is the length of the long arm and L_p is the length of the short arm. A summary of variants by chromosome is provided in Dataset S2.

lower average growth rate of diploids. More work is needed to test these and other possible explanations, and to relate the molecular changes caused by spontaneous mutation to their effects on fitness (e.g., ref. 39).

Discussion

Despite intense study of DNA replication and repair in yeast, the effect that ploidy has on the rate and spectrum of spontaneous, unselected mutations has not been investigated on a genome-wide scale before this study. Populations of haploid and diploid yeast frequently show distinct evolutionary behavior in experimental systems (17, 18, 40–42), but it is unclear what role ploidy-specific mutational variation plays in these differences. We find evidence that multiple dimensions of the mutation spectrum depend on ploidy state in an otherwise isogenic background. In particular, diploidy confers greater replication fidelity with respect to single-nucleotide changes and mt mutations, but diploids are also more prone to large-scale mutations with deleterious fitness effects.

Models of life cycle evolution assume that twice as many mutations occur in diploid genomes than in haploid genomes (e.g., refs. 43, 44, as well as many others), but our findings indicate that this classic and seemingly universal assumption does not hold. Rather, the genomic diploid/haploid mutation rate ratio is only ∼1.4 in the wild-type genetic background we studied. Considering the "mutation load" (the reduction in fitness caused by deleterious mutations in a constant environment), our results, if confirmed in other organisms, imply that diploid life cycles would be favored across a broader range of parameters than suggested by previous theory (as reviewed in ref. 2). On the other hand, diploids would also have less access to beneficial substitutions, reducing their ability to proliferate in a changing environment even more than predicted by previous theory (45).

The higher mutation rate per base pair we observed in haploids appears to be linked to replication timing and could reflect a haploid-specific mechanistic constraint on replication fidelity. It is also possible that our lines carry haploid-specific mutator

alleles, which might be subject to weak purifying selection, given their indirect effects on fitness (46) and the apparent rarity of haploid cells in natural populations of this species (47). Comparing mutation rates in haploid and diploid forms of predominantly haploid organisms (e.g., using nonsporulating *S. pombe*) would be an important next step in differentiating these hypotheses.

While we accounted for rates of cell division and applied similar mutation-calling criteria in haploids and diploids, it is worth considering whether other differences between ploidy levels could have biased our mutation rate comparisons. To examine whether our coverage criteria might differentially affect haploid and diploid mutation rate estimates, we reestimated SNM and indel rates using progressively more stringent cutoffs and found no indication that mutation-calling criteria affected our results (*SI Appendix*, Fig. S11).

Selection might have also influenced our results. In particular, recessive lethal mutations can accumulate in diploids but not in haploids, which could lead us to underestimate the haploid mutation rate. This would not explain the higher SNM rate we observed in haploids, but it could affect our comparisons of indel rates. Using the viability of spores from diploid MA lines, Zhu et al. (5) estimated the rate of recessive lethal mutations as 3.2×10^{-5} per diploid genome per generation, involving SNMs and indels. Applying half of this rate to haploids, we would have missed only approximately five mutations among all of the haploid lines due to recessive lethality (a bias of <1%). Similarly, aneuploidy might not be observed in haploid lines if such changes are highly deleterious. However, aneuploidy for 10 different chromosomes has been observed in haploid MA lines with mutator genotypes (7), and doubling times are similar to the wild type for at least some disomic haploids (48), suggesting that selection does not fully explain the lack of aneuploidy events in haploids.

More generally, mutations might accumulate at a lower rate in haploids than diploids because they are fully exposed to selection and not masked in heterozygous form. However, we found no evidence for selection on SNMs, and a consistent pattern of apparent selection across ploidy levels based on genic versus nongenic indel rates (*SI Appendix*, Fig. S2). While the dearth of genic indels suggests purifying selection, mutations may not always arise at the same rate in genic and nongenic regions (49), complicating this interpretation.

We predicted that the rate of indels might be greater in haploids and *rdh54*Δ diploids owing to their reduced opportunity for homologous DNA repair, which did not prove to be the case (Fig. 2). We have lower statistical power with respect to indels, which are rare relative to SNMs. However, given the mean experiment-wide indel rate observed, we would detect a twofold higher indel rate in haploids ∼89% of the time at the $\alpha = 0.05$ level. Similarly, given the mean diploid indel rate, we would detect a twofold higher indel rate in *rdh54*Δ lines versus *RDH54*$^+$ lines ∼85% of the time.

The lack of a difference in indel accumulation between ploidy levels or any interaction with *RDH54* status suggests that most small indels are the result of processes other than DSB repair, such as replication slippage (50), or that most DSBs are repaired when sister chromatids are available (3, 51). However, our results point to another possible effect of ploidy on DNA repair: The finding that replication timing affects mutation rates more strongly in haploid than diploid genomes suggests that the activity of error-prone DNA repair processes in late-replicating regions [specifically translesion synthesis (27)] is more important in haploids, which may explain why an effect of replication timing is detected in some studies [e.g., in haploids (27)] but not others [e.g., in diploids (5)]. Replication timing may also explain the difference we observed in the spectrum of SNMs between ploidy levels, as this difference increased with replication timing (*SI Appendix*, Fig. S12), with late-replicating regions more prone to A-to-T and C-to-G transversions in haploids. Rather than increasing the rate of indels, the most obvious effect of *rdh54*Δ was to increase the rate of aneuploidy, suggesting that this gene plays a role in mitotic segregation.

LOH in diploids will convert some heterozygous mutant sites into sites homozygous for the ancestral allele, causing us to underestimate diploid point mutation rates. Using maximum likelihood to jointly estimate the rate of LOH and the true mutation rate (*SI Appendix*), we find that the rate of LOH (95% CI) is 7.92 (5.69, 10.65) $\times 10^{-5}$ per base pair per generation and that the observed diploid mutation rates are not substantially downwardly biased by LOH [3% bias; corrected diploid rates: SNMs: 2.97 (2.81, 3.15) $\times 10^{-10}$, indels: 2.09 (1.67, 2.57) $\times 10^{-11}$]. These relative rates of mutation and LOH predict that few heterozygous sites will exist in equilibrium asexual populations (*SI Appendix*).

Haploid and diploid yeast differ in development, gene expression, and physiology. Some of these differences are due to ploidy per se, including differences in cell size and expression profiles (52), while others are due to differences in mating competency as determined by the MAT genotype (53). In our study, we chose to examine diploids with the standard MATa/MATα genotype. Thus, results that we ascribe to ploidy differences may also stem from differences in MAT genotype. Although we did not detect an effect of MATa versus MATα in haploids, it would be valuable to examine the mutational profiles of MATa/MATa and MATα/MATα diploids to clarify the role of MAT heterozygosity versus ploidy per se, especially given that MAT is known to affect repair pathway regulation (3) and competitive fitness (54), although at least some differences in DNA repair appear to be independent of MAT genotype (55). Similarly, haploids experienced less fitness decline than diploids after having accumulated mutations in our experiment, but additional data are needed to determine whether this fitness difference stems from cell size and content effects of ploidy per se, MAT composition, or explanations involving beneficial mutations and dominance.

Previous studies have reported faster rates of adaptation in haploid yeast compared with either MATa/MATα diploids (30, 42) or MATa/MATa diploids (41). Explanations have focused on the greater efficacy of selection in haploids than in diploids because the benefits of mutations are partially masked in diploids, assuming all else is equal. Our observation that haploids experience a higher SNP mutation rate, however, indicates that not all else is equal, and previous observations of more rapid haploid adaptation could, in part, reflect a higher mutation rate.

We observed an order of magnitude higher rate of MA in mt sequences within haploid cells than diploid cells. The number of mt genomes differs from that of nuclear genomes, as multiple mt genomes are transmitted to daughter cells (56). However, mt segregation restores homoplasmy in under ∼20 cell divisions following a mating event between cells with different mt genotypes (56, 57), suggesting that heteroplasmy should have been transient in our 1,600-generation MA experiment. Furthermore, relaxing our filters to include heteroplasmic mt variants had little effect on our results (*SI Appendix*).

We also found that nuclear coverage relative to mt coverage was consistent across ploidy levels ($t = 1.62$, df = 218, $P = 0.11$), suggesting that our diploid lines have about twice the mtDNA of haploids [qualitatively similar results are reported elsewhere (58)]. Assuming sequencing and mapping are equally efficient for nuclear and mtDNA, we estimate that there are three to four sequenced mtDNA genomes per haploid cell, fewer than most previous estimates, although there is substantial variation among strains and conditions (e.g., refs. 56–58). Furthermore, we also see substantial variation in coverage across the mt genome, so coverage metrics may poorly estimate the number of mt genomes per cell.

While we cannot exclude the possibility that a higher effective mt population size in diploids permitted more effective selection

against new mutations, most of the mt mutations we observed were nongenic and presumably had minimal fitness effects. Instead, the factors causing a higher nuclear mutation rate in haploids may have had an even greater impact on mt sequences, or haploids may incur a higher level of mtDNA damage. A study using a reporter gene approach found that mt microsatellites were 100-fold more stable in diploids than in haploids (59), attributable to nuclear ploidy per se rather than *MAT*-specific gene expression or mt copy number (60). The effect of ploidy on mt genome maintenance clearly deserves further study.

In an experimental evolution context, haploid populations of *S. cerevisiae* are sometimes found to attain a diploid state (61–63), but it is unclear to what extent this is due to a high rate of diploidization or a large selective advantage of diploidy. We observed no changes in ploidy, indicating that spontaneous diploidization does not occur at a high rate in the strain and experimental conditions we used. Assuming a rate of diploidization equal to our upper 95% confidence limit, a haploid population would incur >130 nonsynonymous point mutations for every ploidy change event. However, our data also suggest that shifts to diploidy will increase the rate of large structural mutations, consistent with analyses of adaptive genome evolution in haploids and diploids, where diploids are more likely to incur beneficial amplifications or deletions of chromosome segments (17, 18, 63). Polyploid yeast seem to adapt even more rapidly than diploids for the same reason (64). Ploidy shifts, which often seem to lack immediate benefits in yeast (61, 65), may therefore be favored in some environments due to their effects on the mutation spectrum, with our results suggesting that haploids have more access to single-nucleotide changes, while diploids experience more structural variation.

We find that the mutation rate and spectrum differ dramatically between the haploid and diploid forms of a common genetic background. Our results contribute to the growing understanding that the mutation process can vary in response to genetic and genomic context, impacting the ability of organisms to withstand the load of deleterious mutations and to adapt to a changing world.

Materials and Methods

Details on growth rate analyses, phenotyping, bioinformatics, and flow cytometry are provided in *SI Appendix*.

Isogenic haploid and diploid strains were generated starting with a single haploid cell, using single-colony bottlenecks throughout the procedure. The strain SEY6211 (*MAT*a, *ho*, *leu2-3 112*, *ura3-52*, *his3-Δ200*, *trp1-*Δ901, *ade2-101*, *suc2-Δ9*) was obtained from the American Type Culture Collection and induced to switch mating type using a standard plasmid transformation protocol. *RDH54* was deleted (positions 383,065–386,180 on chromosome II), replaced with *KanMX* in cells of each mating type via transformation, and confirmed by Sanger sequencing. *MAT*a and *MAT*α haploids were mated to generate diploids with and without the *RDH54* deletion. These strains were frozen as ancestral controls and plated to single colonies to begin the bottlenecking procedure.

MA was conducted on solid yeast-extract-peptone-dextrose media, supplemented with 40 mg/L adenine sulfate, on 6-cm-diameter plates. Plates were incubated at 30 °C, and plates from the previous day were stored at 4 °C as backup. Petite mt mutations result in very small 24-h colonies, and we used the backup plate if all colonies on a plate were petite, if individual colonies could not be distinguished, or if colonies were absent. Backup plates were required in 0.4% of transfers, with no significant differences among treatments.

Early in the experiment, 13 putatively haploid *MAT*α *RDH54*$^+$ lines were found to be diploid due to mating immediately following the mating type switch during strain construction, leading to a mixed colony of *MAT*α and diploid cells; we continued to propagate these diploids and added 15 *MAT*α *RDH54*$^+$ lines by subdividing five existing lines confirmed to be haploid. We later found that the diploid *RDH54*$^+$ lines generated by controlled crossing, but not those obtained by unintentional mating, were trisomic for chromosome XI; furthermore, the ancestral diploid *RDH54*$^+$ strain (but not *rdh54*Δ) used in phenotype assays carries this trisomy (confirmed by Illumina sequencing).

Concurrent with bottleneck 100, each ancestral control genotype was thawed and plated in the same fashion as the MA lines. The next day, a single colony of each MA line and three replicate colonies from each control plate were transferred to 3 mL of liquid media, grown for 3 d at 30 °C, and frozen in 15% glycerol. These frozen cultures were subsequently used for DNA extraction and sequencing, flow cytometry, and phenotyping.

DNA was extracted from 10-mL saturated cultures using a standard phenol/chloroform method and quantified with fluorometry. DNA (0.4 ng) was used to construct libraries using the Illumina Nextera XT kit and protocol. Libraries were pooled such that diploid samples had twice the concentration of haploid samples (to give equivalent coverage per chromosome) and sequenced in a single Illumina NextSeq lane with paired-end 150-bp reads (average coverage per line before screening: haploid 24.5×, diploid 47.6×). Sequence data are available from the National Center for Biotechnology Information Sequence Read Archive (accession no. SRP139886).

Reads were aligned with Burrows–Wheeler aligner *mem* (66), and mutations in nonrepetitive regions were called using GATK HaplotypeCaller (67), following recommended practices. We screened sites based on multiple metrics and calculated mutation rates accounting for the number of callable sites in each line.

ACKNOWLEDGMENTS. We thank A. Kuzmin for assistance with sequencing and transformations; A. Gerstein, M. Dunham, and M. Whitlock for helpful discussions; A. Johnson for flow cytometry support; C. Landry for providing plasmids; and anonymous reviewers for constructive comments on a previous version of the manuscript. This work was supported by a Banting Postdoctoral Fellowship (to N.P.S.) and Natural Sciences and Engineering Research Council Discovery Grant RGPIN-2016-03711 (to S.P.O.).

1. Otto SP, Gerstein AC (2008) The evolution of haploidy and diploidy. *Curr Biol* 18: R1121–R1124.
2. Mable BK, Otto SP (1998) The evolution of life cycles with haploid and diploid phases. *BioEssays* 20:453–462.
3. Skoneczna A, Kaniak A, Skoneczny M (2015) Genetic instability in budding and fission yeast-sources and mechanisms. *FEMS Microbiol Rev* 39:917–967.
4. Nishant KT, et al. (2010) The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLoS Genet* 6:e1001109-15.
5. Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci USA* 111:E2310–E2318.
6. Lynch M, et al. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105:9272–9277.
7. Serero A, Jubin C, Loeillet S, Legoix-Né P, Nicolas AG (2014) Mutational landscape of yeast mutator strains. *Proc Natl Acad Sci USA* 111:1897–1902.
8. Klein HL (1997) RDH54, a RAD54 homologue in Saccharomyces cerevisiae, is required for mitotic diploid-specific recombination and repair and for meiosis. *Genetics* 147: 1533–1543.
9. Smit A, Hubley R, Green P (2015) RepeatMasker Open-4.0. Available at www.repeatmasker.org/. Accessed March 7, 2018.
10. Schrider DR, Hourmozdi JN, Hahn MW (2011) Pervasive multinucleotide mutational events in eukaryotes. *Curr Biol* 21:1051–1054.
11. Mandegar MA, Otto SP (2007) Mitotic recombination counteracts the benefits of genetic segregation. *Proc Biol Sci* 274:1301–1307.
12. Symington LS, Rothstein R, Lisby M (2014) Mechanisms and regulation of mitotic recombination in Saccharomyces cerevisiae. *Genetics* 198:795–835.
13. Shinohara M, et al. (1997) Characterization of the roles of the Saccharomyces cerevisiae RAD54 gene and a homologue of RAD54, RDH54/TID1, in mitosis and meiosis. *Genetics* 147:1545–1556.
14. Tosato V, Sidari S, Bruschi CV (2013) Bridge-induced chromosome translocation in yeast relies upon a Rad54/Rdh54-dependent, Pol32-independent pathway. *PLoS One* 8:e60926-17.
15. McCulley JL, Petes TD (2010) Chromosome rearrangements and aneuploidy in yeast strains lacking both Tel1p and Mec1p reflect deficiencies in two different mechanisms. *Proc Natl Acad Sci USA* 107:11465–11470.
16. Dujon B (2010) Yeast evolutionary genomics. *Nat Rev Genet* 11:512–524.
17. Zhang H, et al. (2013) Gene copy-number variation in haploid and diploid strains of the yeast Saccharomyces cerevisiae. *Genetics* 193:785–801.
18. Gresham D, et al. (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* 4:e1000303-19.
19. Argueso JL, et al. (2008) Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc Natl Acad Sci USA* 105:11845–11850.
20. Sharp NP, Agrawal AF (2016) Low genetic quality alters key dimensions of the mutational spectrum. *PLoS Biol* 14:e1002419-18.
21. Weiller GF (1994) Frequent site-specific mit- deletions at cryptic exon-intron junctions in the COX1 gene of yeast mtDNA. *Curr Genet* 26:542–545.
22. Anderson JB, Sirjusingh C, Ricker N (2004) Haploidy, diploidy and evolution of antifungal drug resistance in Saccharomyces cerevisiae. *Genetics* 168:1915–1923.
23. Tang Y, Gao X-D, Wang Y, Yuan B-F, Feng Y-Q (2012) Widespread existence of cytosine methylation in yeast DNA measured by gas chromatography/mass spectrometry. *Anal Chem* 84:7249–7255.

24. Capuano F, Mülleder M, Kok R, Blom HJ, Ralser M (2014) Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. *Anal Chem* 86:3697–3702.
25. Farlow A, et al. (2015) The spontaneous mutation rate in the fission yeast Schizosaccharomyces pombe. *Genetics* 201:737–744.
26. Behringer MG, Hall DW (2016) Genome-wide estimates of mutation rates and spectrum in Schizosaccharomyces pombe indicate CpG sites are highly mutagenic despite the absence of DNA methylation. *G3 (Bethesda)* 6:149–160.
27. Lang GI, Murray AW (2011) Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol Evol* 3:799–811.
28. Müller CA, et al. (2014) The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* 42:e3.
29. Magwene PM, et al. (2011) Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* 108: 1987–1992.
30. Marad DA, Buskirk SW, Lang GI (2018) Altered access to beneficial mutations slows adaptation and biases fixed mutations in diploids. *Nat Ecol Evol* 2:882–889.
31. Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N, Keightley PD (2015) Extensive de novo mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. *Genome Res* 25:1739–1749.
32. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. *Nature* 458:342–345.
33. Parada L, Misteli T (2002) Chromosome positioning in the interphase nucleus. *Trends Cell Biol* 12:425–432.
34. Therizols P, Duong T, Dujon B, Zimmer C, Fabre E (2010) Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc Natl Acad Sci USA* 107:2025–2030.
35. Kim S, et al. (2017) The dynamic three-dimensional organization of the diploid yeast genome. *eLife* 6:e23623.
36. Hall DW, Joseph SB (2010) A high frequency of beneficial mutations across multiple fitness components in Saccharomyces cerevisiae. *Genetics* 185:1397–1409.
37. Joseph SB, Hall DW (2004) Spontaneous mutations in diploid Saccharomyces cerevisiae: More beneficial than expected. *Genetics* 168:1817–1825.
38. Gerstein AC (2013) Mutational effects depend on ploidy level: All else is not equal. *Biol Lett* 9:20120614.
39. Kraemer SA, Böndel KB, Ness RW, Keightley PD, Colegrave N (2017) Fitness change in relation to mutation number in spontaneous mutation accumulation lines of Chlamydomonas reinhardtii. *Evolution* 71:2918–2929.
40. Gerstein AC, Otto SP (2009) Ploidy and the causes of genomic evolution. *J Hered* 100: 571–581.
41. Gerstein AC, Cleathero LA, Mandegar MA, Otto SP (2011) Haploids adapt faster than diploids across a range of environments. *J Evol Biol* 24:531–540.
42. Zeyl C, Vanderford T, Carter M (2003) An evolutionary advantage of haploidy in large yeast populations. *Science* 299:555–558.
43. Crow JF, Kimura M (1965) Evolution in sexual and asexual populations. *Am Nat* 99: 439–450.
44. Kondrashov AS, Crow JF (1991) Haploidy or diploidy: Which is better? *Nature* 351: 314–315.
45. Orr HA, Otto SP (1994) Does diploidy increase the rate of adaptation? *Genetics* 136: 1475–1480.
46. Lynch M (2010) Evolution of the mutation rate. *Trends Genet* 26:345–352.
47. Ezov TK, et al. (2006) Molecular-genetic biodiversity in a natural population of the yeast Saccharomyces cerevisiae from "Evolution Canyon": Microsatellite polymorphism, ploidy and controversial sexual status. *Genetics* 174:1455–1468.
48. Torres EM, et al. (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317:916–924.
49. Frigola J, et al. (2017) Reduced mutation rate in exons due to differential mismatch repair. *Nat Genet* 49:1684–1692.
50. Fortune JM, et al. (2005) Saccharomyces cerevisiae DNA polymerase δ: High fidelity for base substitutions but lower fidelity for single- and multi-base deletions. *J Biol Chem* 280:29980–29987.
51. Kadyk LC, Hartwell LH (1992) Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. *Genetics* 132: 387–402.
52. Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR (1999) Ploidy regulation of gene expression. *Science* 285:251–254.
53. de Godoy LMF, et al. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455:1251–1254.
54. Selk E, Wills C (1998) Mismatch repair and the accumulation of deleterious mutations influence the competitive advantage of MAT (mating type) heterozygosity in the yeast Saccharomyces cerevisiae. *Genet Res* 71:1–10.
55. Li XC, Tye BK (2011) Ploidy dictates repair pathway choice under DNA replication stress. *Genetics* 187:1031–1040.
56. Birky CW, Strausberg RL, Forster JL, Perlman PS (1978) Vegetative segregation of mitochondria in yeast: Estimating parameters using a random model. *Mol Gen Genet* 158:251–261.
57. Chen XJ, Butow RA (2005) The organization and inheritance of the mitochondrial genome. *Nat Rev Genet* 6:815–825.
58. Grimes GW, Mahler HR, Perlman RS (1974) Nuclear gene dosage effects on mitochondrial mass and DNA. *J Cell Biol* 61:565–574.
59. Sia EA, et al. (2000) Analysis of microsatellite mutations in the mitochondrial DNA of Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* 97:250–255.
60. Sia RAL, Urbonas BL, Sia EA (2003) Effects of ploidy, growth conditions and the mitochondrial nucleoid-associated protein Ilv5p on the rate of mutation of mitochondrial DNA in Saccharomyces cerevisiae. *Curr Genet* 44:26–37.
61. McDonald MJ, Hsieh Y-Y, Yu Y-H, Chang S-L, Leu J-Y (2012) The evolution of low mutation rates in experimental mutator populations of Saccharomyces cerevisiae. *Curr Biol* 22:1235–1240.
62. Gerstein AC, Chun HJ, Grant A, Otto SP (2006) Genomic convergence toward diploidy in Saccharomyces cerevisiae. *PLoS Genet* 2:e145.
63. Fisher KJ, Buskirk SW, Vignogna RC, Marad DA, Lang GI (2018) Cause and consequences of genome duplication in haploid yeast populations. bioRxiv:247320.
64. Selmecki AM, et al. (2015) Polyploidy can drive rapid adaptation in yeast. *Nature* 519: 349–352.
65. Gerstein AC, Otto SP (2011) Cryptic fitness advantage: Diploids invade haploid populations despite lacking any apparent advantage as measured by standard fitness assays. *PLoS One* 6:e26599-13.
66. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
67. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.

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

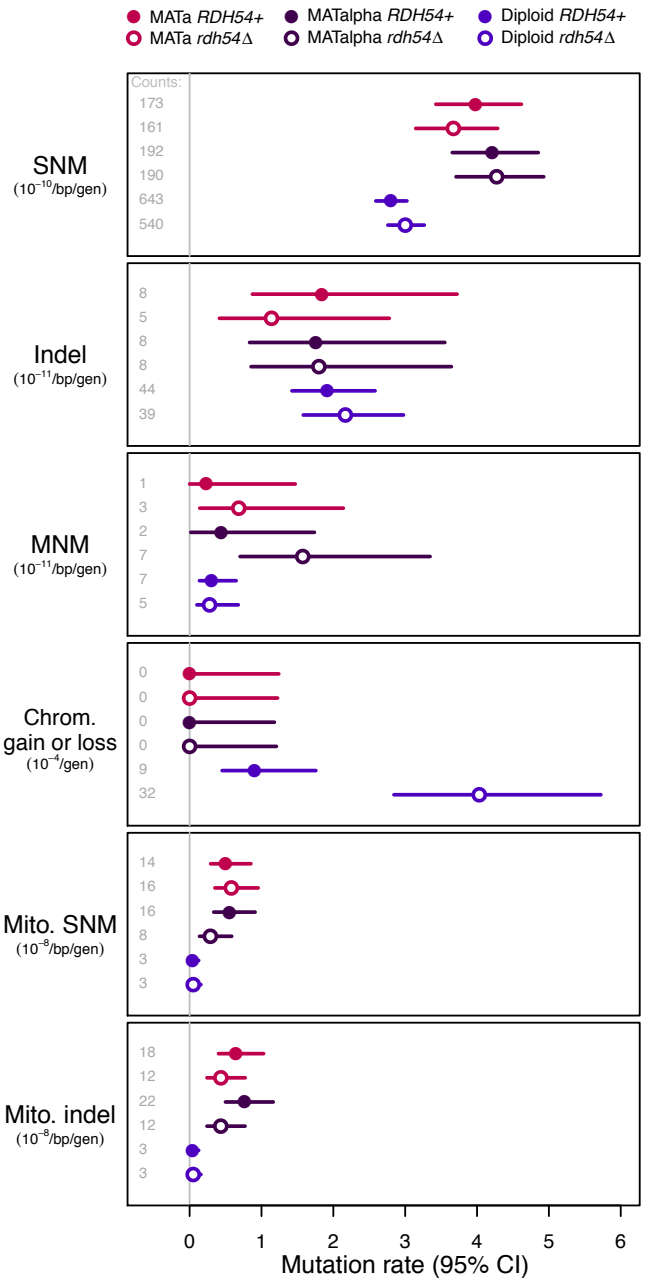

Figure S1. Mutation rates in each group of MA lines, with 95% confidence intervals, showing *MATa* and *MATα* separately. Mutation categories are as in Fig. 2. Mating type within haploids did not significantly affect the rate of any mutation type.

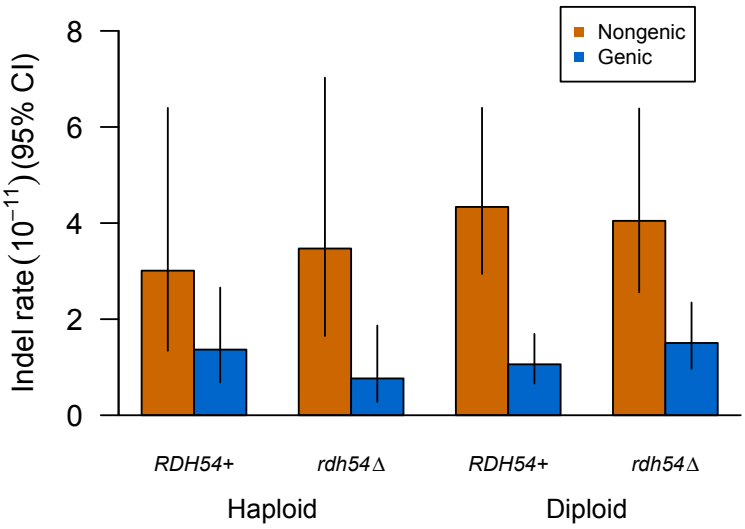

Figure S2. Rates of indel mutation per base pair per generation in each treatment for genic and non-genic regions. Rates of genic indels did not differ among groups (G = 2.56, df = 3, P = 0.47); rates of non-genic indels did not differ among groups (G = 0.93, df = 3, P = 0.82).

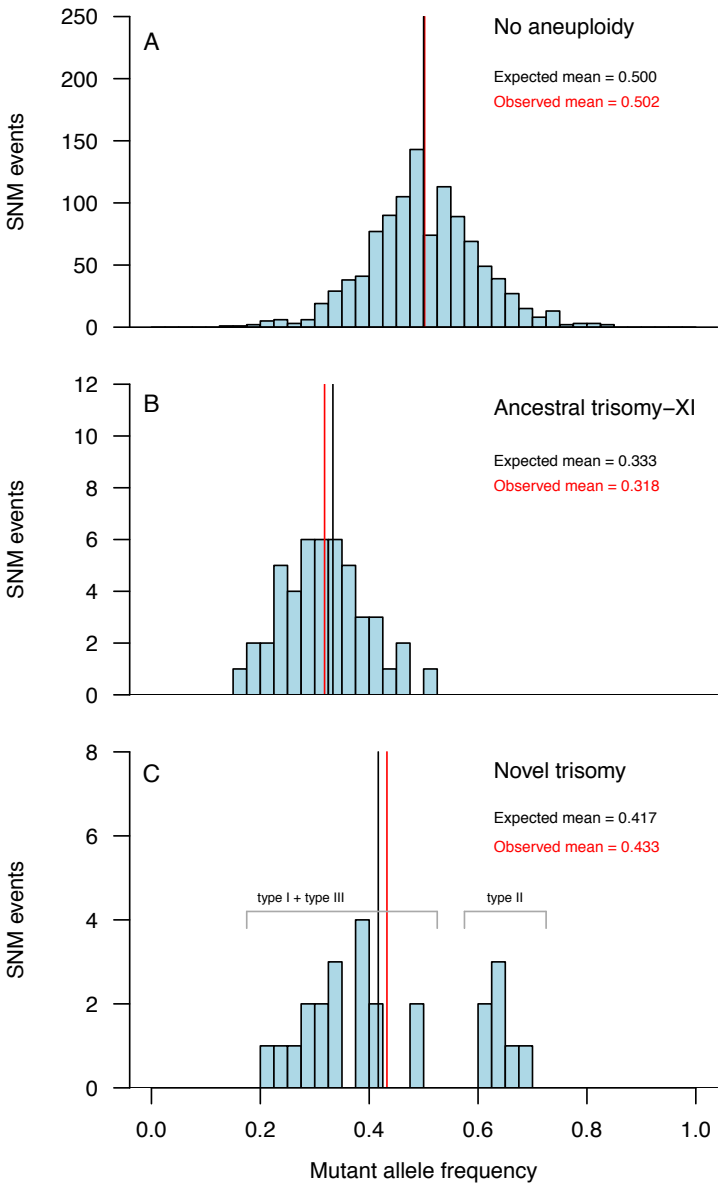

Figure S3. Allele frequencies inferred from depth of coverage for diploid heterozygous SNMs, accounting for different histories of aneuploidy. (A) Most events occurred on non-aneuploid chromosomes, where the observed mean coverage frequency for the mutant allele did not differ from the 0.5 expectation (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).

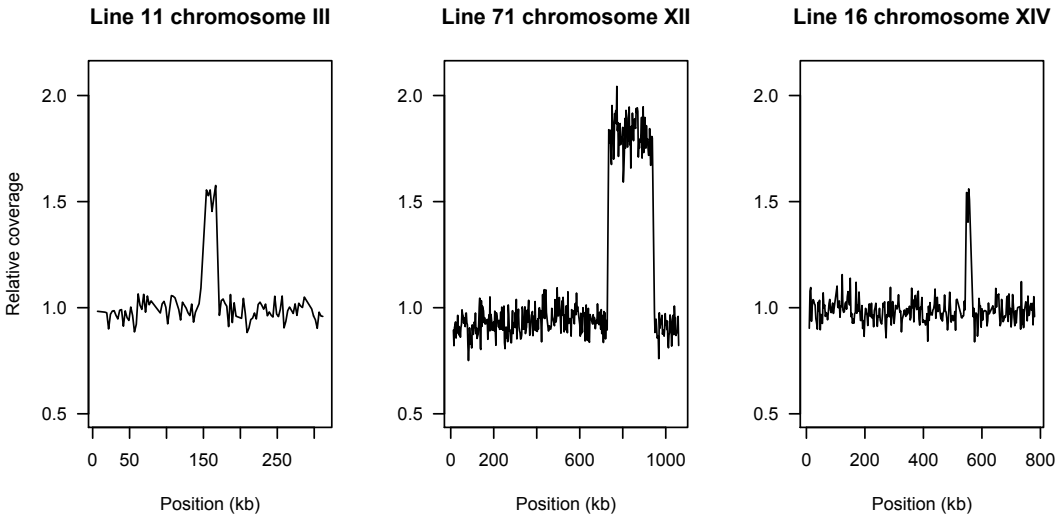

Figure S4. Large segmental duplications. Relative coverage in 2.5 kb windows for chromosomes and samples with suspected duplications. Here relative coverage represents the coverage ratio for a given window (mean coverage for sites in that window divided by median coverage for that sample and chromosome) 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.

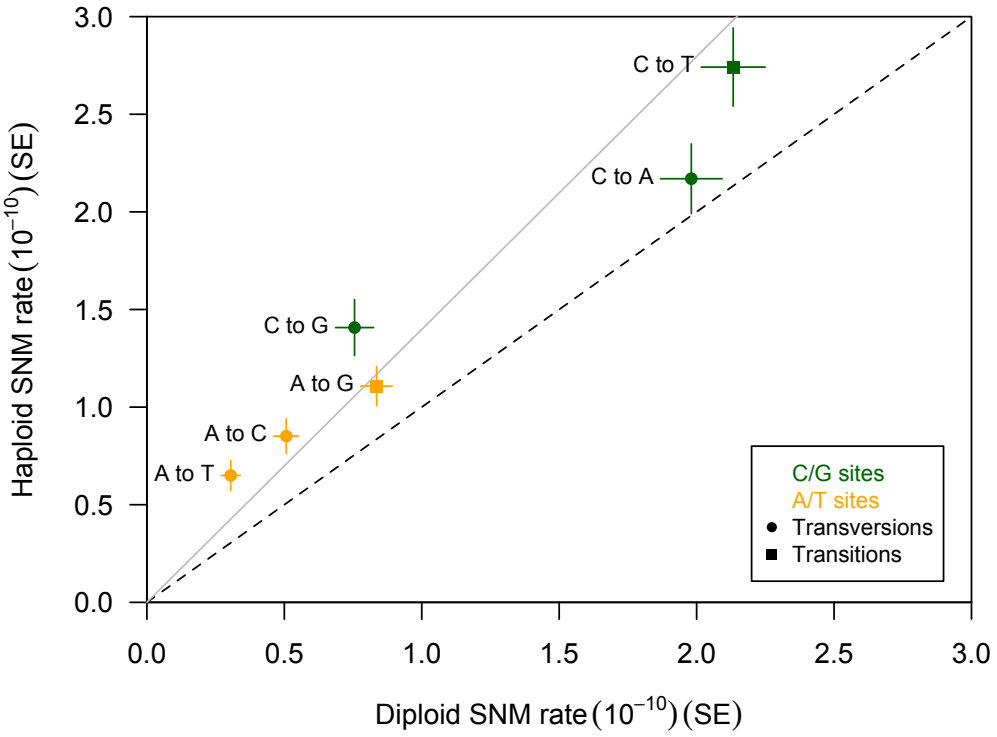

Figure S5. Haploid and diploid SNM rates per base pair per generation for each of the six mutation types, including the complementary change. The dashed line shows a 1:1 relationship; the solid gray line shows the slope corresponding to the overall haploid:diploid SNM rate ratio. Haploids had a higher 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).

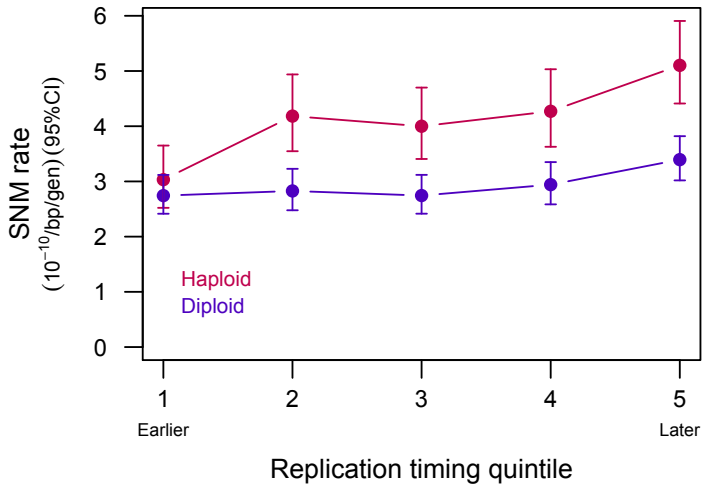

Figure S6. SNM rates in relation to replication time, for haploids and diploids. The genome was divided into five roughly-equal bins based on replication timing, and SNM rate calculated for sites in each bin. Haploids and diploids have a similar SNM rate in early-replicating sites (quintile 1), and haploids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

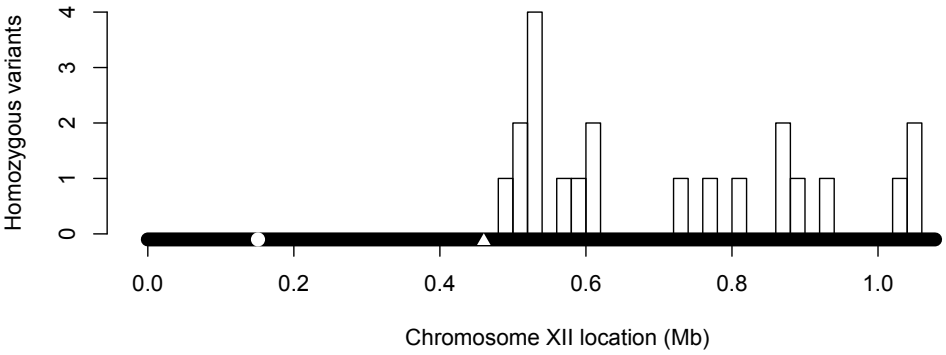

Figure S7. Locations of homozygous variants on chromosome XII relative to the centromere (white circle) and the region of rDNA repeats (white triangle). All homozygous variants on this chromosome were distal to the rDNA repeat region, and all occurred in unique MA lines. See Dataset S2 for homozygous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

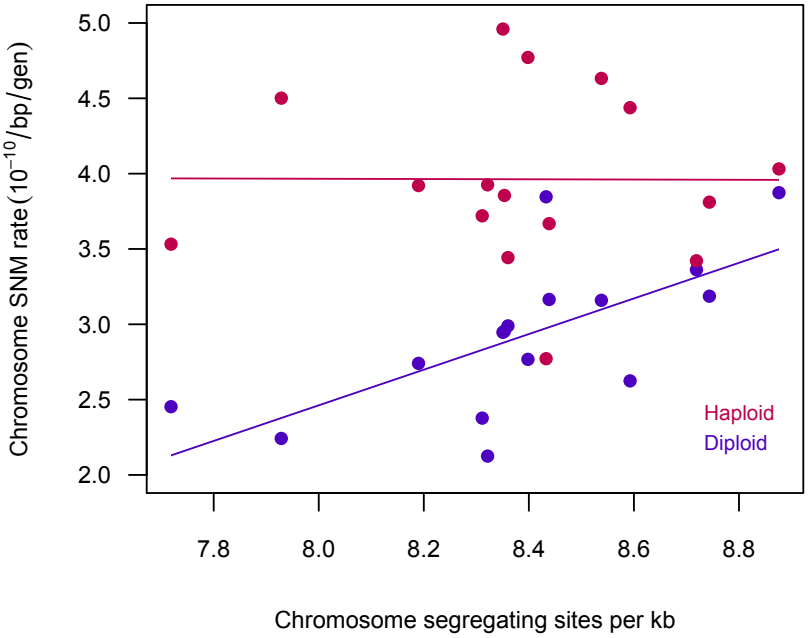

Figure S8. SNM rate on each chromosome in haploids or diploids versus number of polymorphic sites on that chromosome estimated from population samples. Polymorphism was correlated with SNM rate in diploids ($\rho = 0.79$, $P < 0.001$), but not haploids ($\rho = -0.05$, $P = 0.87$).

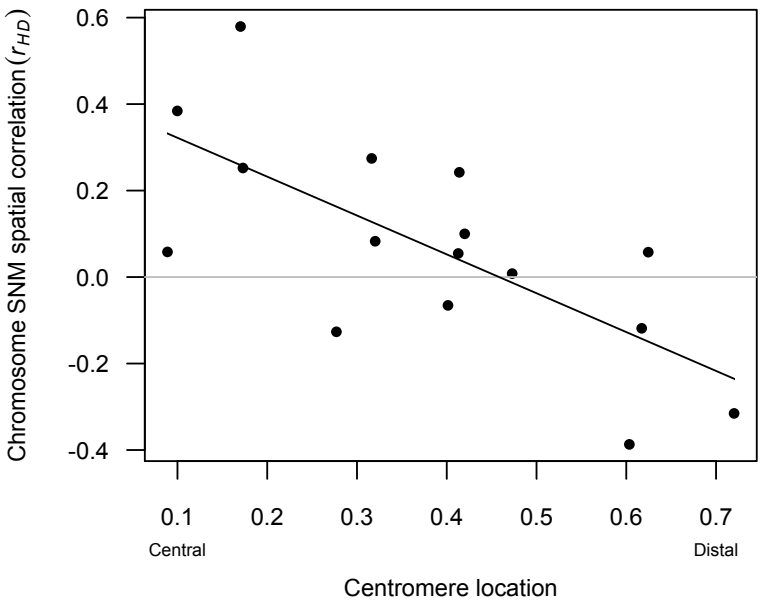

Figure S9. Spatial correlation of haploid versus diploid SNMs, r_{HD}, on each chromosome, plotted against centromere location relative to the chromosome center, $(L_q - L_p)/ (L_q + L_p)$. We calculated r_{HD} by estimating Gaussian kernel density of SNMs at 512 equally-spaced points along the chromosome 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, $P < 0.01$).

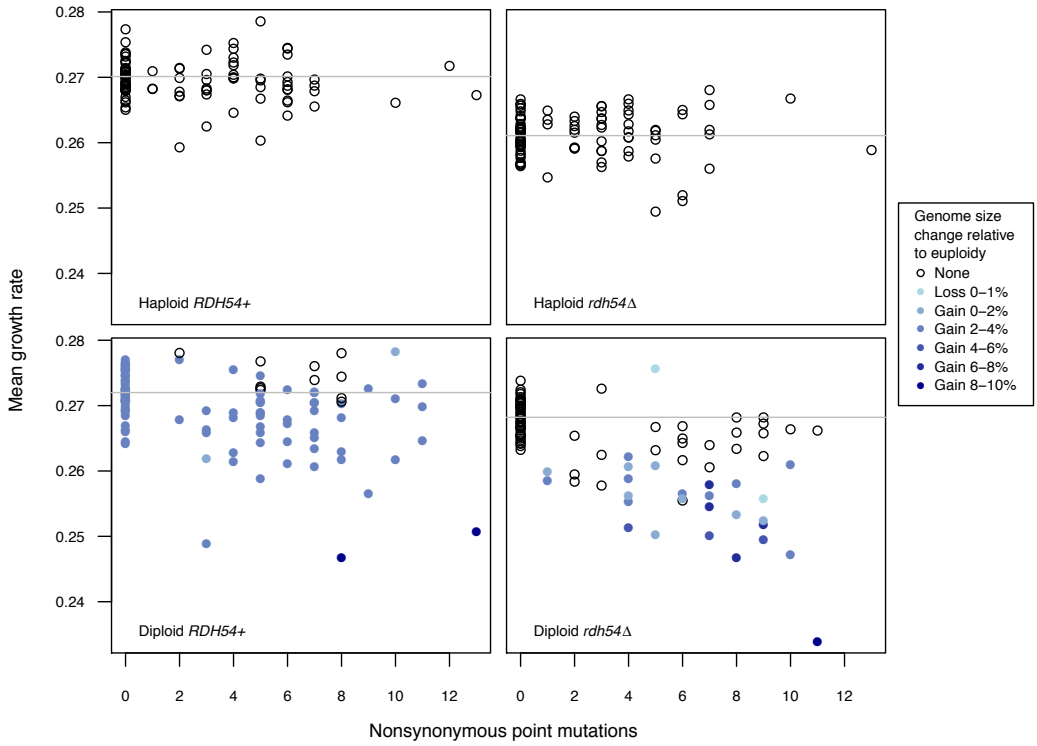

Figure S10. Mean growth rate (maximum rate of OD increase per hour) plotted against number of non-synonymous point mutations for each line. Colour codes indicate genome size change relative to euploidy. Ancestral control lines are plotted and assigned a value of zero for non-synonymous point mutations (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.

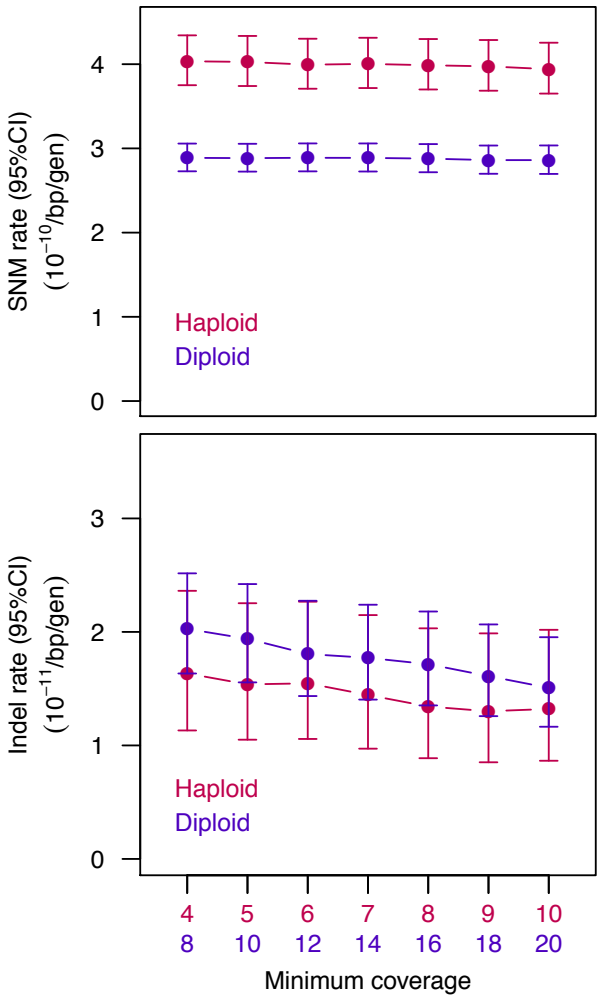

Figure S11. Mutation rate estimates using different minimum coverage cutoffs for SNMs (top) and indels (bottom). For each minimum coverage value, the number of mutations and callable sites was scored in haploids and diploids, where minimum diploid coverage equals two times minimum haploid coverage. Our 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.

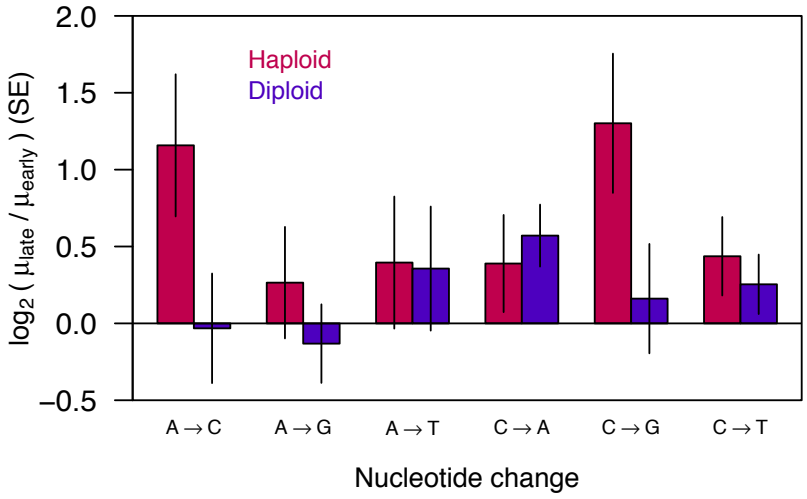

Figure S12. Changes in the rate of each type of SNM with replication timing. Bars indicate the log ratio of substitution rate for sites in the latest quintile of replication timing (μ_{late}) relative to the rate for sites in the earliest quintile (μ_{early}), accounting for detection power by site type 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 $P < 0.01$).