




Upper bound on the mutational burden imposed by a CRISPR-Cas9 gene-drive element

Michael S. Overton ^{1,†} Sean E. Guy,^{1,3,†} Xingsen Chen,^{1,4} Alena Martsul,¹ Krypton Carolino,^{1,5} Omar S. Akbari ², Justin R. Meyer,¹ Sergey Kryazhimskiy ^{1,*}

¹Department of Ecology, Behavior and Evolution, University of California San Diego, La Jolla, CA 92093, United States

²Department of Cell and Developmental Biology, University of California San Diego, La Jolla, CA 92093, United States

³Present affiliation: Department of Anatomy, Biochemistry & Physiology, University of Hawaii at Manoa, Honolulu, HI 96822, United States

⁴Present affiliation: Department of Entomology, University of Arizona, Tucson, AZ 85721, United States

⁵Present affiliation: Becton Dickinson Biosciences, San Diego, CA 92130, United States

*Corresponding author: Department of Ecology, Behavior and Evolution, University of California San Diego, La Jolla, CA 92093, United States. Email: skryazhi@ucsd.edu

[†]These authors contributed equally.

Homing-based CRISPR-Cas9 gene drives (CCGDs) are powerful tools for genetic control of wild populations, with applications from disease eradication to species conservation. However, Cas9 alone and in a complex with guide RNA can cause double-stranded DNA breaks at off-target sites, which could increase the mutational load and lead to unintended loss-of-heterozygosity (LOH) events. These undesired effects raise potential concerns about the long-term evolutionary safety of CCGDs, but the magnitude of these effects is unknown. To measure how the presence of a CCGD or a Cas9 alone in the genome affects the rates of LOH events and de novo mutations, we carried out a mutation accumulation experiment in yeast *Saccharomyces cerevisiae*. We found no detectable effects on the genome-wide rates of mutations or LOH events. Our power calculations suggest that CCGD or Cas9 affect these rates by less than 30%, which is much less than natural variation for these traits in yeast. A more detailed examination shows that CCGD or Cas9 may alter the lengths and genomic distributions of LOH events, but the statistical support for these effects is weak. Thus, our results demonstrate that CCGDs impose at most a weak additional mutational burden in the yeast model. Although mutagenic effects of gene drives need to be further evaluated in other systems, our results add credence to the proposition that the evolutionary risks posed by well-designed gene drives may be acceptable.

Keywords: gene drive; mutations; loss-of-heterozygosity; off-target effects

Introduction

Homing-based CRISPR-Cas9 gene drives (CCGDs) are synthetic genetic elements that can rapidly spread in sexual populations through super-Mendelian inheritance (Bier 2022). The ability to encode various functions within CCGDs promises to give us an unprecedented degree of control over wild populations (Champer et al. 2016). For example, by encoding a *Plasmodium*-disrupting peptide within a CCGD, it may be possible to reduce or even eliminate the spread of malaria (Wang and Jacobs-Lorena 2013; Gantz et al. 2015; Adolphi et al. 2020); a CCGD encoding drought resistance may allow us to prevent the extinction of populations vulnerable to climate change (Li et al. 2008; Champer et al. 2016). However, large-scale deployment of CCGDs in the wild faces several significant biological, as well as ethical, challenges (Esvelt et al. 2014; Oberhofer et al. 2018; Rode et al. 2019). Two major biological challenges are caused by evolution (Rode et al. 2019). The first, short-term, problem is that certain mutations can arise in the CCGD element itself, abolishing its activity. Given that CCGD carriage often comes with a fitness cost, such loss-of-function mutations would be favored by selection and would result in gene drive-resistant populations (Champer et al. 2017; Unckless et al. 2017; Beaghton et al. 2019). This problem is widely recognized, and various engineering solutions have been proposed to mitigate

it (Marshall et al. 2017; Marshall and Akbari 2018; Kandul et al. 2021). The second problem is that Cas9 has off-target effects (Cradick et al. 2013; Hsu et al. 2013; Pattanayak et al. 2013; Lin et al. 2014), which could increase the incidence of new mutations as well as loss-of-heterozygosity (LOH) events (Smukowski Heil 2023) across the genome, potentially altering long-term evolutionary trajectories of treated populations. However, we know very little about the magnitude of these effects and their evolutionary significance (Rode et al. 2019; Garrood et al. 2021). Here, we set out to address this latter gap.

CCGDs vary in their design and complexity, but every CCGD contains a guide RNA (gRNA) that targets a particular genomic sequence (“the target allele”) and a Cas9 endonuclease (Gantz and Bier 2015; Champer et al. 2016). In addition, CCGDs can also contain a genetic “payload” that produces a desired phenotype. The drive construct must be flanked by sequences homologous to those surrounding the target allele. When an engineered organism homozygous for the CCGD mates with a wildtype individual carrying the target allele, the fusion of gametes brings the drive and the target sequences together into the same cell. The Cas9/gRNA complex cuts the target allele, and the entire CCGD can be copied in its place by the homology-directed repair mechanism, thereby converting an initially heterozygous offspring into one homozygous for the CCGD allele. Through this process, CCGDs can achieve up

to 98% inheritance, rapidly fixing the desired phenotype in the target population (Gantz et al. 2015; Adolphi et al. 2020).

The presence of a CCGD in the genome can change the rates and types of genome-wide mutations via several known as well as possibly other, as of yet unknown, mechanisms (Cradick et al. 2013; Fu et al. 2013; Hsu et al. 2013; Pattanayak et al. 2013; Lin et al. 2014; Tsai et al. 2017; Newton et al. 2019). The best understood mechanism is template promiscuity whereby the Cas9/gRNA complex binds and cuts DNA sequences that are similar but not identical to the target (Fu et al. 2013; Hsu et al. 2013; Lin et al. 2014; Boyle et al. 2021). These off-target dsDNA cuts are repaired by the homology-directed repair or non-homologous end-joining, leading to LOH and indel mutations, respectively, at loci other than the target (Cradick et al. 2013; Fu et al. 2013). Off-target mutations may also be generated because the Cas9-gRNA complex bound to DNA might interfere with normal replisome progression. Importantly, such binding transiently occurs not only at the sequences similar to the target but also at random PAM sites (Sternberg et al. 2014). Moreover, the Cas9 protein alone has a high, nonspecific affinity for DNA (Sternberg et al. 2014), which can potentially introduce additional mutations at non-PAM associated sites through the same interference mechanism. These additional mutations and LOH events could have unpredictable evolutionary consequences. For example, some LOH events could resolve hybrid incompatibilities and improve fitness (Mandegar and Otto 2007; Smukowski Heil et al. 2017; James et al. 2019). Perhaps more likely, by reducing genetic diversity and exposing recessive deleterious alleles, LOH events could potentially lead to population decline, particularly in species that are already endangered (Frankham 1995; Wernberg et al. 2018). However, the severity of these long-term evolutionary consequences depends on how strongly CCGDs affect the rates of mutations and LOH events, on the type of these events, and on their distribution along the genome. Thus, to quantify evolutionary risks associated with CCGDs, we need to measure how CCGDs affect the rates and the distributions of these events.

To address this problem, we designed a mutation accumulation (MA) experiment in yeast *Saccharomyces cerevisiae* with the aim to detect potential effects of a CCGD or Cas9 alone on the rates and genomic distributions of LOH events and single-nucleotide mutations (SNMs). In contrast to some other methods for detecting off-target effects (Tsai et al. 2015, 2017; Akcakaya et al. 2018; Allen et al. 2019), MA experiments allow us to detect mutagenic effects caused by any molecular mechanism, including unknown ones. One major advantage of using yeast—an established and powerful system for testing CCGDs (DiCarlo et al. 2015; Roggenkamp et al. 2018)—is that we can easily detect LOH events by carrying out MA experiments in “hybrid” strains that are heterozygous at many sites throughout the genome (Loeillet et al. 2020; Pankajam et al. 2020; Sui et al. 2020; Dutta et al. 2021; Vijayan et al. 2025). Another advantage is that we can maintain hundreds of independent MA lines for hundreds of generations (Yin and Petes 2013; Zhu et al. 2014; Zheng et al. 2016; Dutta et al. 2017, 2021; Sharp et al. 2018; Loeillet et al. 2020; Pankajam et al. 2020; Sui et al. 2020), which gives us substantial statistical power to detect even relatively small CCGD- or Cas9-induced effects. The main focus of this paper is on LOH events because they result from the same molecular mechanism (homology-directed repair) that supports the intended activity of the CCGD and because they occur at high rates in yeast which further increases our statistical power to detect CCGD- and Cas9-induced effects.

Materials and methods

Experimental methods

Strain construction

All strains are listed in [Supplementary Table 1 in File 1](#). We generated the W, C, and D strains by mating two haploid yeast strains. One parent strain is derived from the laboratory *S. cerevisiae* strain BY4741 (BY; MAT α his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0 ho) and the other parent is derived from the vineyard strain YAN501 (RM; MAT α his3 Δ 1 leu2 Δ 0 ura3 Δ 0 ho::KanMX) generously provided by Alex Nguyen Ba (U Toronto; Nguyen Ba et al. 2022). These two parent strains differ by ~40,000 SNPs and small indels (Bloom et al. 2013). We used standard transformation techniques to replace the his3 Δ 1 allele in BY parent with a functional version to facilitate diploid selection (Gietz and Schiestl 2007). We amplified the NatMX-GFP, NatMX-GFP-Cas9, and NatMX-GFP-Cas9-gRNA constructs from a previously constructed gene drive cassette, p414-Cas9, and p426-gRNA[ADE2] (DiCarlo et al. 2013, 2015) and integrated each of them into each of the parent strains at the ADE2 locus (Chr XV:564476 to 566191) using standard yeast transformation methods and selection on yeast extract, peptone, dextrose (YPD) + nourseothricin (nat) plates. The gRNA sequence used here targets the 5' end of the ADE2 gene, and the disruption of ADE2 produces an easily observed red colony phenotype (DiCarlo et al. 2015). Single colonies of BY and RM haploid parents containing each of the integrated constructs were mated, diploids were selected on G418 + his $^-$ plates, and correct diploid progeny were confirmed by PCR targeting the ADE2 region. Diploid founders were each derived from independently transformed and mated BY and RM parents to reduce the chance of multiple lines being established from a mutated founder. To test the activity of our CCGD, we also constructed two MAT α “tester” haploids derived from strain MJM64 generously provided by Michael McDonald (Monash U; McDonald et al. 2016) by knocking in one of two functional copies of ADE2. MJM64 was transformed with the BY4742 copy of ADE2 to generate the “intact target” strain, yKC57, and with the YAC2485 (DiCarlo et al. 2015) copy of ADE2 sequence to generate the “no target” strain yKC29. The YAC2485 ADE2 sequence is engineered with 12 mismatches to abolish gRNA targeting. In both tester strains, URA3 is driven by the haploid-specific promoter STEP5pr which makes them 5-FOA sensitive as haploids but not as diploids. They also carry a KanMX cassette, which makes their progeny G418 resistant.

Mutation accumulation experiment

For each of the three constructs, we used eight founders to establish 95 independent MA lines, for a total of 285 MA lines. Each of these lines was independently propagated for 43 cycles of streaking to single colony isolation on YPD agar plates and incubating at 30 °C for 48 h, for a total average of 800 generations. The colony closest to a premarked location on the plate was used to continue the next cycle to prevent any potential artificial selection effects. Founder and end-point clones were preserved by culturing single colonies in YPD liquid media to mid-log growth and freezing at –80 °C in 15% glycerol. To estimate the number of elapsed generations, we grew one founder and 10 end-point clones from each of the W, C, and D backgrounds under identical conditions to the MA experiment in four biological replicates. For each replicate, we picked a colony, resuspended it in PBS, estimated the cell count using the Coulter Counter, and calculated the number of generations within a cycle as log₂ (colony size).

Testing of gene-drive activity

To test whether the gene drive elements remained active by the end of the MA experiment, we sampled six end-point D clones, three exhibiting low LOH accumulation and three chosen randomly. We sporulated and dissected tetrads for each diploid clone and selected *MAT α* haploids that were *ura*⁻ and *KanMX*⁻, making them 5-fluoroorotic acid (5FOA) resistant and G418 sensitive. All mating types of all haploids were confirmed with creep assays (Arras et al. 2022) against strains MJM36 (*MAT α*) and MJM64 (*MAT α*) (McDonald et al. 2016). Since 5-FOA and G418 require different media pH, we first selected clones on Complete Synthetic Media (CSM) + 5FOA pH 4.5, then on CSM + G418 pH 7.0. For each original D clone, we picked one *MAT α* haploid and mated it with the two *MAT α* “tester” haploids. After mating, we select for diploids using 5-FOA and G418 selection, as detailed above. After 48 h of growth on CSM + G418, colonies were photographed. The presence of red colonies indicates successful *ade2* disruption by an active CCGD (DiCarlo et al. 2013, 2015).

Library preparation and sequencing

Genomic DNA was extracted and purified from each founder and end-point clone using a modified ethanol precipitation protocol. Briefly, overnight cultures were pelleted and resuspended in 6% SENT (6% SDS, 10 mM EDTA, 30 mM Tris, pH 8) and incubated at 65 °C for 15 min. Cooled lysates were combined with RNase A and incubated at 37 °C for 60 min. A half volume of 3 M NaOAc was added and the mixture was centrifuged to pellet proteins and cellular components. The supernatant was washed first in isopropanol and then ice-cold ethanol. DNA pellets were air dried in inverted tubes and resuspended in nuclease-free water. We prepared whole-genome sequencing libraries using the Illumina Nextera system with a modified protocol (Baym et al. 2015). We pooled 100 bp paired-end libraries and sequenced them on an Illumina Hi-Seq 4000 platform to a median read depth of 34.2 \times .

Gene drive biosafety

All work involving gene drive constructs was performed according to protocols approved by the Institutional Biosafety Committee at UCSD. Specifically, great care was taken to prevent aerosolization and escape by culturing only on solid media using disposable spreaders and plates and housed in secondary containers. All work areas were decontaminated with 10% bleach. A split drive system (Cas9 gene and gRNAs on a separate drug-selected plasmids) was used when possible.

Data analysis

Analysis of statistical power

Theoretically predicted power. To determine the statistical power of our MA experiment to detect differences in the rates of LOH events between strains, we modeled their accumulation as a Poisson process with rate $\lambda = 1.76 \times 10^{-2}$ per genome per generation (Pankajam et al. 2020; Sui et al. 2020). Then, assuming that our experiment would take about $T = 750$ generations, the number of LOH events in each end-point clone is a Poisson random number with expectation $\mu = \lambda T = 13.2$. Given that we had the resources to maintain the maximum of 95 MA lines per founder, we simulated our experiment as follows. We first formed the “baseline” sample by drawing 95 random numbers from the Poisson distribution with mean $\mu = 13.2$. We then generated the “alternate” sample of end-point clones whose event counts were drawn from the Poisson distribution with mean $\mu' = \mu(1 + f_e)$ where $f_e > 0$ is the effect size. We then compared the differences in the means of these two samples using a

permutation test with 1,000 independent permutations, and called the difference significant if the P-value was at or below 0.01. For each effect size f_e between 5% and 50% with the step size of 5%, we repeated this procedure 1,000 times and estimated the power of our experiment as the fraction of simulations that yielded a significant test. We estimate the theoretically predicted power to detect changes in the SNM rates analogously, using a Poisson process with rate $\lambda = 4.0 \times 10^{-2}$ per genome per generation (Zhu et al. 2014).

Observed power estimation. We estimate the actual statistical power of our experiment as follows. We first pool all 229 of our end-point clones that carry a total of $n = 1,645$ LOH events to form the “baseline” sampling population. Then, at each realization of our simulation for a given effect size f_e , we form the “alternate” population by randomly distributing $n(1 + f_e)$ additional LOH events (rounded to the nearest integer) across all 229 clones. We then randomly sample 79 clones from the baseline population and 67 clones from the alternate population and compare their sample means using the same permutation test as above. We carry out 1,000 realizations at each effect size f_e between 5% and 50% with the step-size of 5%, and obtain an estimate of power as above. The observed statistical power for point mutations was estimated analogously. We estimate the actual power to detect changes in the SNM rates analogously, using a total of $n = 336$ observed SNMs/indels.

Estimating the expected number of mutations in the construct

Since our MA lines accumulated mutations for approximately 800 generations, it is possible that some mutations occurred in the construct and some of them may have deactivated the gene drive element. SNP and indels occur on average at rates of 1.67×10^{-10} and 7.5×10^{-12} per base pair per generation, respectively (Lang and Murray 2008; Zhu et al. 2014; Pankajam et al. 2020). Given that our constructs D, C, and W are 7,999, 7,601, and 2,770 bp long, respectively, a typical end-point clone from each respective strain is expected to have 2.23×10^{-3} , 2.12×10^{-3} , and 7.73×10^{-4} mutations in the construct. Thus, we expect a total of 0.15, 0.18, and 0.061 D, C, and W end-point clones across the entire experiment to carry a mutation in the construct. Given that the length of the gRNA and Cas9 genes are 388 and 4,890 bp (including promoter and terminator), respectively, we can conservatively estimate the fractions of gene-drive/Cas9 deactivating mutations as 66.0% and 64.3% in the D and C clones, respectively. Thus, we expect 0.090 and 0.11 of the D and C clones to have a gene-drive or Cas9 deactivating mutation, respectively.

These estimates are based on the reasonable assumption that all mutations within the constructs are effectively neutral in our MA experiment, i.e. they do not provide a selective advantage or disadvantage substantial enough to overcome the strong genetic drift imposed by periodic single-cell bottlenecks. A more important source of uncertainty in these estimates arise from the fact that they are based on genome-wide average mutation rate estimates, but the actual mutation rates can vary up to 6-fold across locations in the genome (Lang and Murray 2011). Thus, we might reasonably expect anywhere between 0.065 and 2.34 out of 229 sequenced clones to carry a mutation in the construct. Similarly, we expect between 0.015 and 0.54 of the D clones to have a gene-drive-deactivating mutation and between 0.018 and 0.66 of the C clones to have a Cas9-deactivating mutation.

Analysis of sequencing data

Sequencing library preparation failed for five founders (H_F00, H_H00, N_F00, N_H00, F_D00). In order to use their descendant

MA lines for LOH and mutation analyses, we imputed the ancestral genotypes from descendant end-point clones, as described below. Inclusion of these lines did not change our results. In addition, we found that the ancestors of two D MA lines (2 founders and 27 end-point clones) were apparently triploid as detailed in Overton and Kryazhimskiy (2025), compromising our ability to detect new LOH events, and were excluded. We also excluded 19 clones from further analysis due to low coverage or traces of cross-contamination during library preparation.

Genotyping. We genotyped our founder and end-point clones at the initially heterozygous marker sites using our newly developed reference-symmetric genotyping pipeline (Overton and Kryazhimskiy 2025). Briefly, we align trimmed reads to both the BY and RM parental reference genomes. We exclude 1,774,809 bp of repeat regions identified by RepeatMasker (Smit et al. 2023) and 7.5 kb at chromosome ends (Nguyen Ba et al. 2022), as all of these regions are difficult to map unambiguously. This reduces the length of the genome that we monitor for the presence of LOH events to an effective length of 10.3 Mb (20.6 Mb for SNMs), and this length is used for all genome wide per base-pair rate calculations. To obtain genotype calls, we first use the GATK HaplotypeCaller (McKenna et al. 2010; DePristo et al. 2011; Poplin et al. 2017; van der Auwera and O'Connor 2020) with the BY and the RM reference genomes separately, reconcile the emitted calls at each marker site, and then remove dubious heterozygous calls.

For each of the four founders whose genomes were not sequenced (see above), we impute the genotype at a marker site as heterozygous (homozygous for a given allele) if at least six of its descendant end-point clones and all other founders genotyped at that site are called heterozygous (homozygous for the same allele).

Detection and exclusion of aneuploidies. We detect aneuploidy events as described by Overton and Kryazhimskiy (2025). We found no aneuploid events in the founders, but detected a total of 21 aneuploidies (12 chromosome losses and 9 chromosome gains) in 17 end-point clones, with 5, 8, and 6 events affecting the W, C, and D strains, respectively. All aneuploid chromosomes were excluded from LOH analyses.

Identification of LOH events. To identify LOH events, we first find the so-called “LOH tracts” in an end-point clone as sequences of adjacent marker sites that are called homozygous for the same allele. A boundary of an LOH tract is estimated as the midpoint between its first converted marker and the adjacent unconverted marker (or a marker converted to a different homolog). We merge multiple LOH tracts into a single LOH event if the LOH-tract boundaries are separated by less than 10 kb (Sui et al. 2020). We classify those LOH events that contain the last marker on a chromosome arm as terminal (tLOH); other LOH events are classified as interstitial (iLOH). The set of detected LOH events is listed in Supplementary Table 2 in File 1.

Corrections for undetected LOH events. After detecting LOH events and identifying their lengths, we apply three corrections for undetected iLOH and tLOH events: (i) correction for inter-marker iLOH events; (ii) correction for iLOH events overlapping with tLOH events; (iii) correction for overlapping tLOH events. These corrections are described in detail in Overton and Kryazhimskiy (2025). We compare both detected (i.e. uncorrected) and corrected LOH counts across strains, as indicated.

Estimation of causal dsDNA breakpoint positions. We estimate the positions of dsDNA breaks that likely led to the formation of an LOH event as follows. For iLOH events and tLOH events under 20 kb, we locate the breakpoint as the midpoint between its boundaries (Hum and Jinks-Robertson 2017). For tLOH events over 20 kb, we locate the breakpoint inside the tLOH event, 10 kb from its centromere-proximal boundary (Sui et al. 2020). However, we note that this estimate is not very precise because the degree of strand resection—and therefore the length of the LOH region extending from the breakpoint—differs between the two dsDNA repair mechanisms (strand crossover during double-strand break repair or break induced-replication repair) that can produce tLOH events (Malkova et al. 2001; Symington et al. 2014).

Detection of de novo mutations. To detect de novo SNMs and indels we start with the set of genotypes emitted by our genotyping pipeline and exclude all sites that differ between the two parental reference sequences or sites that we identified as ancestrally heterozygous. The remaining variants form the set of putative mutations. This set contains many sites with multiple identical base substitutions in individual clones. Given that the mutation rate in yeast is about 1.67×10^{-10} (Zhu et al. 2014), the expected number of mutations that occur at a given site during the entire MA experiment in any of the clones is $\mu = 1.67 \times 10^{-10} \times 800 \times 229 = 3.06 \times 10^{-5}$. Therefore, the probability that two or more mutations occur at a given site is $P_{2+} = 1 - e^{-\mu}(1 + \mu) \approx \mu^2/2 = 4.68 \times 10^{-10}$, so that we would expect $2LP_{2+} = 2 \times 1.2 \times 10^7 \times 4.68 \times 10^{-10} = 1.12 \times 10^{-2}$ sites across the whole genome to be mutated twice or more in the entire experiment, where $L = 1.2 \times 10^7$ is the haploid genome length of *S. cerevisiae*. Since indel rates are generally much lower than those of SNMs, about 5.0×10^{-12} per base pair per generation (Zhu et al. 2014), it is extremely unlikely that any duplicate SNMs or indels have occurred in our experiment. Therefore, we identify only unique variants (those that are present only in a single end-point clone) as new mutations. We identified 349 of such unique variants. 46 of them form 23 sets, such that mutations within each set are located within 100 bp of each other in the same clone, an event that is unlikely to happen by chance. Thus, we consider each such set as one “complex” mutation event (Makarova and Burgers 2015; Sui et al. 2020). The final set of detected mutations is listed in Supplementary Table 3 in File 1.

Over the course of the MA experiment, a de novo mutation present initially as a heterozygote may be overwritten by an LOH event and not detected. We correct for this detection bias in our mutation rates using the approach by Sharp et al. (2018). Specifically, we compute $\mu_{\text{corrected}} = [N_{\text{HET}}(1 + \rho/2) + N_{\text{ALT}}]/T/C$, where N_{HET} is the number of heterozygous mutations, ρ is the average fraction of the genome under LOH, N_{ALT} is the number of mutations converted to the homozygous nonreference genotype, T is the number of generations, and C is the number of clones.

Comparison of LOH rates across strains

We define the LOH event rate as the number of LOH events per unit time (irrespective of their length), and we define the LOH conversion rate as the probability per unit time that a position will be converted from heterozygous to homozygous state by an LOH event (Overton and Kryazhimskiy 2025). We estimate genome-wide and chromosome-arm-specific LOH event rates and LOH conversion rates as in Overton and Kryazhimskiy (2025). To analyze finer-scale rate variation, we divide each chromosome into 50 kb non-overlapping windows as described by Overton and Kryazhimskiy (2025). (Note that Figs. 3, Supplementary 6 and 7 in File 2 show results for overlapping 50 kb windows, which

produces smoother profiles, but all statistical tests are done for non-overlapping windows.) We then calculate the per-basepair per generation event rate for each window for strain s as $\lambda_s = k_s/n_s \cdot TL$, where k_s is the number of LOH events detected in that window across all MA lines descendent from founder s corrected for undetected events (see above), n_s is the number of these MA lines, $T = 800$ is number of generations, and L is the window size (50 kb for most windows). To estimate the 95% confidence intervals within each window and strain, we use the bootstrapping approach, i.e. we resample the number of events with replacement from their empirical distribution across clones.

To test whether the distributions of LOH events across the genome are statistically different between two founders, we calculate the difference in their cumulative distributions of LOH events across each chromosome using the Wasserstein statistic from the two samples R package (Dowd 2020, 2023). This statistic is calculated as the integral of the function $|E_1(x) - E_2(x)|$, where E_1 and E_2 are the empirical cumulative distribution functions for the two strains (Vaserstein 1969). P -values are calculated from 5,000 iterations of a Monte Carlo simulation where we randomly reshuffle the founder labels of all LOH events.

To test for differences in genome-wide LOH conversion rates between two founders, we perform a permutation test on the mean fraction of the genome converted by reshuffling founder labels across LOH events. To test for differences in local LOH conversion rates, we perform permutation tests of the mean rate on each 50 kb window by reshuffling founder labels across clones. We then apply Benjamini-Hochberg correction with $FDR \leq 0.05$ across windows for each founder comparison.

gRNA similarity. The probability that the Cas9:gRNA complex binds and cuts DNA at a given locus depends on the presence of a PAM site and on the 20-bp sequence immediately upstream of it (Hsu et al. 2013; Sternberg et al. 2014; Boyle et al. 2021). To identify genomic regions where the cutting activity of the Cas9:gRNA complex might be elevated, we first index all genomic positions with either the canonical -NGG or the weakly interacting -NAG PAM sequences on both strands of the BY and RM reference genomes. Then, for each PAM position, we score the 20-bp region immediately upstream against our gRNA sequence using the empirically-informed scoring function getMITScore from the package crisprScore in R (Hsu et al. 2013; Hoberecht et al. 2022). To look for regions that may have elevated LOH rate in founder $s = C$ or D compared with the W founder, we divided the genome into non-overlapping 50 kb windows and computed the normalized LOH event rate in window i as $r_{si} = (R_{si} - R_{Wi}) / (R_{si} + R_{Wi})$, where R_{si} and R_{Wi} are the LOH event rates in window i in founder s and W , respectively. We correlate this normalized rate with the maximum gRNA similarity score in each window (using the mean gRNA similarity score produces qualitatively similar results).

Results

Mutation accumulation experiment for detecting the effects of a CCGD on LOH and SNM rates

To measure the genome-wide rates of LOH events and SNMs in the presence and absence of a CCGD element, we designed a MA experiment using a hybrid-based scheme similar to that implemented in several recent studies (Loeillet et al. 2020; Pankajam et al. 2020; Sui et al. 2020; Dutta et al. 2021; Vijayan et al. 2025). The hybrid ancestors of our MA experiment (later referred to as “founders”) were generated by crossing the laboratory strain BY and the vineyard isolate RM (see section “Strain construction” in

Materials and methods and Supplementary Table 1 in File 1). The two parent strains differ from each other by approximately 40,000 SNPs and small indels spaced on average every 320 bp apart across the yeast genome (Bloom et al. 2013), such that the founders are heterozygous at these “marker” sites. An LOH event that occurs during the MA experiment can be detected if it converts one or more adjacent initially heterozygous markers to one of the parental homologs (Fig. 1a). While we cannot detect very short LOH events or LOH events that occur in repetitive genome regions that are difficult to map (see section “Analysis of sequencing data” in Materials and methods and Overton and Kryazhimskiy 2025), most of heterozygosity losses in yeast occur through large so-called terminal LOH events that typically extend over tens to hundreds of thousands of basepairs (Yin and Petes 2013; Zheng et al. 2016; Pankajam et al. 2020; Dutta et al. 2021). Therefore, the relatively high density of markers in our founders allows us to detect the vast majority of LOH events that contribute to heterozygosity losses in yeast.

We constructed three types of founder hybrids by integrating different constructs into the *ADE2* locus (Fig. 1b, Supplementary

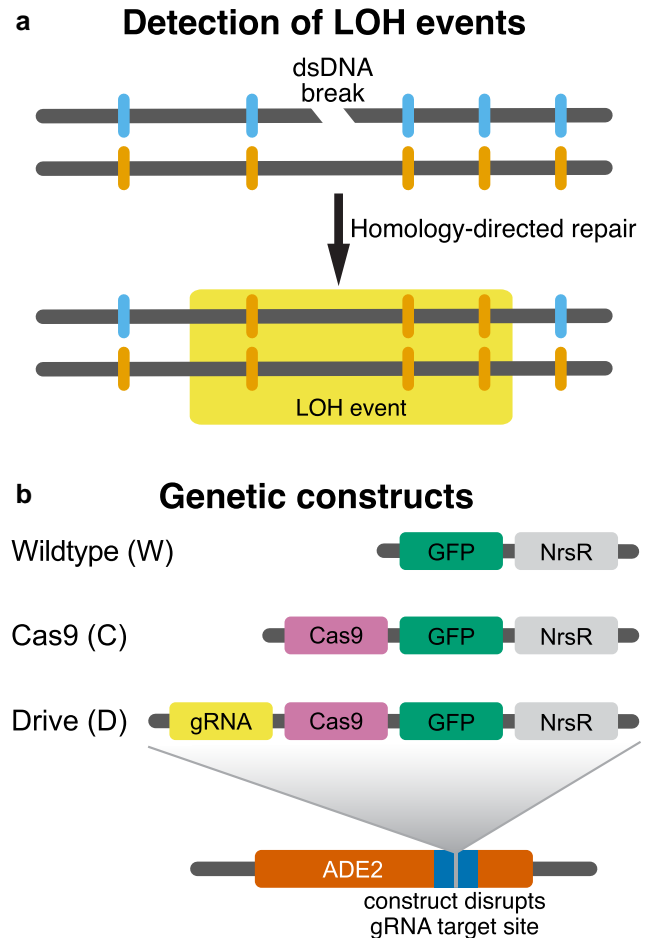


Fig. 1. Schematic of the experimental design. (a) Detection of LOH events. Our founder strains are diploid hybrids heterozygous at about 40,000 marker sites across the genome (blue and orange notches on top and bottom strands, respectively). If a dsDNA break is repaired by homology-directed repair, it causes loss-of-heterozygosity at nearby marker sites, which can be detected by whole-genome sequencing. (b) Design of genetic constructs for the three founder strains. NrsR is the nourseothricin N-acetyl transferase gene which confers resistance to nourseothricin. All constructs are integrated into the *ADE2* locus at Chr XV and disrupt the gRNA binding site within this gene.

Table 1 in File 1; see Section “Strain construction” in Materials and methods). The “drive” founders (or “D” for short) are homozygous for the constitutively expressed CCGD with the gRNA targeting *ADE2* (Fig. 1b). Since the *ADE2* gene is disrupted by the integrated CCGD cassette, the target sequence is absent from the genome of D strains, such that any cut performed by the Cas9/gRNA complex would be off-target. This type of strain was designed to simulate a gene drive that had successfully fixed in a population, and is now permanently integrated into the species’ genome. The “Cas9” founders (or “C” for short) are homozygous for the constitutively expressed Cas9 gene. We designed this type of strain to look for possible mutagenic effects of a “naked” Cas9 protein, i.e. one that is not associated with any gRNAs, something that can occur for example if the Cas9 and gRNA expression levels are not perfectly matched (Hsu et al. 2013). Finally, the wildtype control founders (or “W” for short) are homozygous for the integration cassette without any gene-drive components.

Our a priori expectation was that the main potential side-effect of the presence of a CCGD element in the genome would be an increase in the rate of LOH events at off-target sites driven by the homology-directed repair mechanism. Previous studies reported up to about 50-fold variation in LOH event rates among yeast hybrids (Dutta et al. 2017, 2021; Pankajam et al. 2020; Sui et al. 2020). Since we anticipated that the CCGD-associated effects would likely be smaller, we sought to design a sufficiently powerful MA experiment to detect less than 2-fold changes in LOH rates. Modeling the accumulation of LOH events in each MA line as a Poisson process we found that, if each wildtype MA line accumulates on average 13 LOH events, the probability of detecting a 15% increase in the LOH rate with 95 MA lines per founder at P-value of 0.01 would be 85.7% (Fig. 2a; see section “Analysis of statistical power” in the Materials and methods). Given that the previously estimated rates of LOH events average to 2.3×10^{-2} (range from 3.6×10^{-3} to 4.7×10^{-2}) per genome per generation (Dutta et al. 2017, 2021; Tattini et al. 2019; Loeillet et al. 2020; Pankajam et al. 2020; Sui et al. 2020; Tutaj et al. 2022), we expected to achieve this statistical power by propagating our MA lines for 750 cell divisions. We also carried out a similar power analysis for detecting SNMs, which we discuss below in section “Upper bound on the

increases in the genome-wide mutation rates imposed by CCGD or Cas9”.

Guided by these a priori estimates, we isolated 8, 8, and 7 founder clones of types D, C, and W, respectively, and established between 11 and 14 MA lines per clone, which resulted in 95 MA lines per D, C, and W type each, for a total of 285 MA lines. We propagated these lines using a standard MA protocol (Halligan and Keightley 2009) (see section “Mutation accumulation experiment” in the Materials and methods). Briefly, for each MA line, we picked a random colony from a previous growth cycle, streaked it to single cells and let the resulting colonies grow for 48 h. Assuming that each growth cycle corresponds to 20 cell divisions, we predicted that a typical MA line would accumulate 13 LOH events after 38 cycles. To account for variation among LOH-rate estimates across studies, we propagated our MA lines for 43 cycles. After the experiment was completed, we counted the number of cells per colony after 48 h of growth and estimated that our MA lines in fact underwent approximately 800 cell divisions (see Section “Mutation accumulation experiment” in Materials and methods). We then sequenced the full genomes of all founder and end-point clones to a median depth of 34.2x. After excluding clones that did not pass our quality-control checks (see Section “Analysis of sequencing data” in Materials and methods), we retained 79, 83, and 67 end-point clones for subsequent analyses from W, C, and D lines, respectively.

Before estimating LOH rates from these data, we used two approaches to verify that the CCGD element remained active during the experiment. First, we picked six end-point D clones and experimentally confirmed that their CCGD element was in fact active (see Section “Testing gene-drive activity” in Materials and methods and Supplementary Fig. 1a in File 2). Second, we used our whole-genome sequencing data to identify any de novo mutations that occurred within our constructs and could have disrupted the CCGD function. We found a total of six end-point clones (one W, three C, and two D clones) with such mutations, with each clone carrying either a single heterozygous SNP or indel within the construct (Supplementary Fig. 1b in File 2). Only two of these mutations occurred in the drive components, a non-synonymous SNP (A299T) in the Cas9 gene of a C clone and an

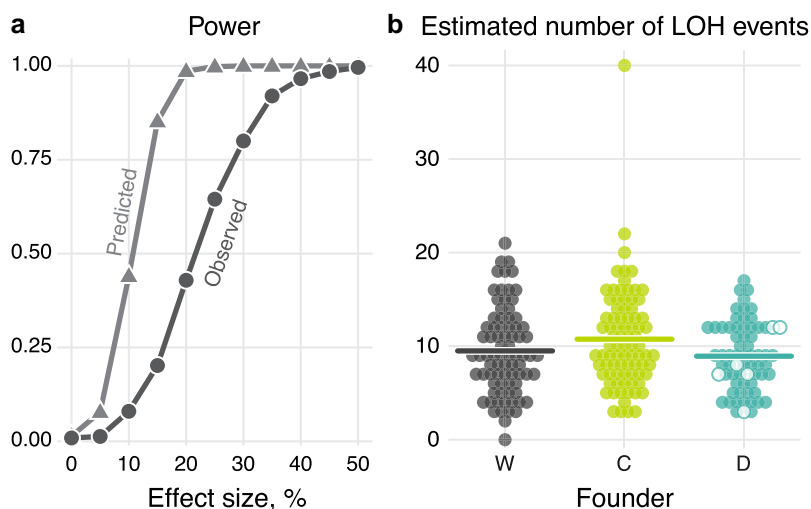


Fig. 2. Statistical power to detect differences in the LOH rates and the estimated number of LOH events in our mutation accumulation experiment. (a) The theoretically predicted and the observed power to detect a certain percentage increase in the LOH rate at P-value 0.01 (see text for details). (b) The number of the LOH events per end-point clone of each founder after correcting for undetected events. Hollow circles indicate clones tested for CCGD activity (see Supplementary Fig. 1a in File 2 and Section “Testing of gene-drive activity” in the Materials and methods). Horizontal lines indicate the mean number of LOH events per clone.

indel at the 3' end of the gRNA promoter in a D clone. The number of observed construct mutations is roughly consistent with our expectation (see Section “[Estimating the expected number of mutations in the construct](#)” in Materials and methods). We conclude that even if all of these mutations deactivated the gene-drive components, they would only minimally decrease our power to detect differences in LOH rates between founders.

In summary, we designed and carried out an MA experiment with a reasonable statistical power to detect possible effects of a gene drive and/or naked Cas9 on the rates of LOH events. All available evidence strongly suggests that the gene-drive components in our D and C MA lines remained active during this experiment.

Upper bound on the increase in the genome-wide LOH-rate imposed by CCGD or Cas9

We next used a novel dual-reference pipeline that we developed ([Overton and Kryazhimskiy 2025](#)) to genotype our end-point and founder clones and identify LOH events. We discovered a total of 1,645 LOH events across all of our 229 our MA lines (a median of 7 per end-point clone), supported by a total of 279,788 converted markers ([Supplementary Table 2 in File 1](#)). We detected on average $648/83 = 7.8$ and $444/67 = 6.6$ LOH events per C and D line, respectively ([Table 1, Supplementary Fig. 2 in File 2](#)), which is statistically indistinguishable from $553/79 = 7.0$ events detected in a typical W line ($P = 0.131$ and $P = 0.419$ for C and D lines, respectively; permutation test).

After correcting for undetected events (see Section “[Analysis of sequencing data](#)” in the Materials and methods and [Overton and Kryazhimskiy 2025](#)), we estimate that a total of 2,348 LOH events with length exceeding 17 bp likely occurred across our 229 MA lines. We estimate that a typical C and D line accumulated $936/83 = 11.2$ and $627/67 = 9.35$ of LOH events, respectively ([Fig. 2b](#) and [Table 1](#)), statistically indistinguishable from $785/79 = 9.94$ events accumulated by a typical W line ($P = 0.131$ and $P = 0.419$ for C and D lines, respectively; permutation test).

On average, 5.5% of the genome per clone lost heterozygosity during our MA experiment. This fraction was 4.9% and 5.4% for a typical C and D clone, respectively, statistically indistinguishable from 6.1% for a typical W clone ($P = 0.205$ and $P = 0.652$, permutation test; [Table 1, Supplementary Fig. 3 in File 2](#)).

Given the absence of a detectable effect of the CCGD or Cas9 on the overall LOH event rates, we sought to estimate the upper bounds on these effects given the power of our experiment. Since the estimated numbers of LOH events that occurred during our experiment are somewhat lower than predicted theoretically and have greater variance than a Poisson distribution ([Supplementary Fig. 4 in File 2](#)), the statistical power of our experiment is slightly reduced. When we re-estimate the power of our experiment, the chance of detecting a 15% LOH-rate increase at the P -value 0.01 in our data is in fact about 20% instead of the anticipated 85.2% ([Fig. 2a](#); see Section “[Analysis of statistical power](#)” in Materials and methods for details). Nevertheless, our experiment gives us power close to 100% to detect the effect sizes above 45% and substantial power to detect the effect size above 30%. Therefore, our results indicate that the presence of a CCGD or the Cas9 protein alone increases the LOH rates in yeast (if at all) likely by less than 30% and almost certainly by less than 45%.

Effects of CCGD and Cas9 on LOH characteristics and their genomic distribution

Given that we detected over 1,600 LOH events, we have enough statistical power to examine not only the potential CCGD- or

Cas9-induced effects on the genome-wide rate of LOH event but also test whether CCGD or Cas9 alone have more subtle effects.

Interstitial versus terminal LOH events

Double-stranded DNA breaks can be repaired by multiple homology-directed repair mechanisms ([Malkova et al. 1996, 2001](#); [Mehta and Haber 2014](#); [Hum and Jinks-Robertson 2017](#)), resulting in the so-called “interstitial” and “terminal” LOH events (iLOH and tLOH). In particular, tLOH events are associated with replication stalling during S-phase and may occur due to replisome collisions with protein-DNA complexes ([Labib and Hodgson 2007](#); [Shyian and Shore 2021](#)). Since the DNA-binding activity of Cas9, alone or in complex with a gRNA, could interfere with the DNA processing during homology-directed repair or produce replication fork stalling, it is possible that the presence of these elements could change the rates of iLOH and tLOH events differently even if the total LOH rate changes are undetectable. To test for this possibility, we classified all 2,348 inferred LOH events into 1,898 (80.8%) interstitial and 450 (19.2%) terminal events (see Section “[Analysis of sequencing data](#)” in Materials and methods). We found that the C and D strains accumulated on average 1.84 and 2.04 tLOH events, respectively, which is statistically indistinguishable from 2.03 tLOH events accumulated by the W strain ($P = 0.389$ and $P = 0.950$, respectively, permutation test; [Table 1](#)). The C and D clones accumulated on average 9.44 and 7.31 iLOH events, which again were statistically indistinguishable from the 7.91 events found in a typical W clone ($P = 0.056$ and $P = 0.373$, respectively), though the proportion of interstitial events in the C strain is 83.8%, which is marginally higher than 79.6% in the W strain ($P = 0.035$; χ^2 test).

LOH length distributions

Interactions between Cas9, with or without a gRNA, and genomic DNA could also disrupt key repair processes such as strand resection, strand invasion, or polymerase progression, which could potentially lead to changes in the length distributions of ensuing LOH events. We find that the iLOH length distributions are bimodal in all strains ([Supplementary Fig. 5 in File 2](#)), possibly due to the fact that different repair mechanisms tend to produce events of different lengths ([Symington et al. 2014](#)), but the median length of an iLOH event is statistically indistinguishable between strains (471, 430, and 752 bp in the W, C, and D backgrounds; $P = 0.544$ and $P = 0.207$ for C vs W and D vs W comparison; permutation test; see Section “[Comparison of LOH rates across strains](#)” in Materials and methods). Terminal events, on the other hand, are somewhat shorter in both C (median = 131 kb, $P = 0.034$) and D (median = 142 kb, $P = 0.133$) strains than in the W control (median = 213 kb).

Distribution of LOH events across the genome

Given that the CCGD off-target activity is sequence specific, it is possible that CCGD or the naked Cas9 protein strongly affect LOH rates only in certain regions of the genome and that such local effects are missed when the data are pooled across the whole genome. Thus, we next looked for possible effects of these elements at the chromosome-arm and 50-kb scales, which trade-off statistical power with resolution. We found no difference in the distributions of tLOH events across strains at either scale ([Supplementary Fig. 6 in File 2](#)). At the same time, while the distributions of iLOH events are indistinguishable between strains at the 50-kb scale ($P > 0.074$, Wasserstein test after Benjamini-Hochberg correction; [Fig. 3b](#)), they are somewhat different at the chromosome-arm scale ($P = 0.047$ and 0.036 for C and D strains compared with W; χ^2 test; [Fig. 3a](#)). However, we could not localize these differences to any specific chromosome arms ($P > 0.192$, permutation test after

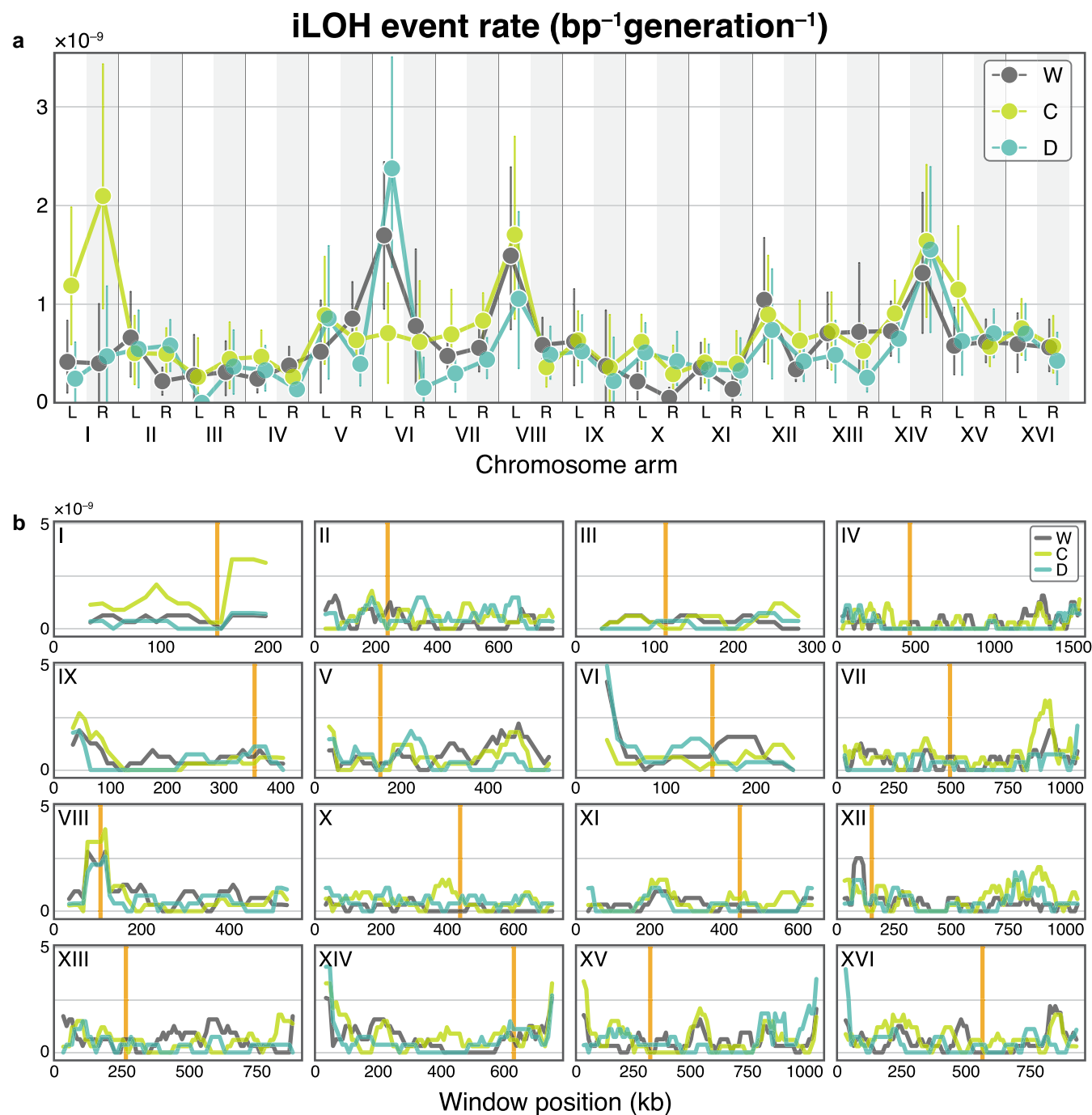


Fig. 3. Distribution of iLOH event rates across the genome. (a) iLOH event rates across chromosome arms. Vertical bars indicate the 95% bootstrap confidence intervals. (b) iLOH event rates across 50 kb sliding windows on each chromosome with a 10 kb step size. Orange vertical lines indicate centromeres.

Benjamini–Hochberg correction), nor did we find any detectable effect of this difference on the distribution of LOH conversions across the genome (Supplementary Fig. 7 in File 2).

We also tested whether LOH event rate variation is associated with certain nucleotide sequences. To this end, we scanned the BY and RM genomes for PAM sites and scored the similarity of the 5′-adjacent sequences to our gRNA (see Section “Comparison of LOH rates across strains” in Materials and methods). We found no evidence that the LOH event rates were elevated in C or D backgrounds in those 50 kb windows that harbored more motifs similar to our gRNA (Supplementary Fig. 8a in File 2) or those where

densities of PAM sites were higher ($P > 0.558$, linear regression; Supplementary Fig. 8b, c in File 2).

Taken together, these results suggest that, to the first approximation, the LOH characteristics and their distribution along the genome are largely unaffected by the presence of a CCGD or naked Cas9. However, the full picture is more nuanced. We carried out several statistical comparisons and found four with P -values slightly below 0.05, with the four being the effect of Cas9 (i) on the proportion of iLOH events and (ii) on the lengths of tLOH events, and the effects of (iii) CCGD and (iv) Cas9 on the genome distribution of iLOH events. Given that these results are only

Table 1. Key statistics and estimates.

End-point clones	Loss-of-heterozygosity					Single-nucleotide mutations					Indels		
	Detected	Estimated ^a	Event rate ^b , $\times 10^{-2}$	Conv. rate ^c , $\times 10^{-5}$	Estimated iLOH	Detected	Estimated ^d	Rate ^e , $\times 10^{-11}$	Detected	Estimated	Rate ^e , $\times 10^{-11}$		
W	553	785	1.24 ± 0.13	7.61 ± 1.36	625	70	71.9	5.49 ± 1.46	48	49.3	3.76 ± 0.99		
C	648	936	1.40 ± 0.16	6.11 ± 1.03	784	77	79.1	5.74 ± 1.53	43	44.2	3.21 ± 1.17		
D	444	627	1.16 ± 0.12	6.73 ± 1.26	490	53	54.5	4.90 ± 1.27	45	46.2	4.16 ± 1.18		

^a After correction for undetected events.

^b Per genome per generation ± 95% bootstrap confidence intervals.

^c Per base-pair per generation ± 95% bootstrap confidence intervals.

^d After correction for events overwritten by LOH events.

marginally significant, we cannot rule out the possibility that they represent Type I errors. But the possibility that CCGD and Cas9 elements slightly change some LOH characteristics and their distribution also remains.

Upper bound on the increases in the genome-wide mutation rates imposed by CCGD or Cas9

In addition to affecting LOH rates, the presence of the CCGD or Cas9 protein could also increase the rate of new mutations. To test this, we identified new mutations in our end-point clones (Supplementary Table 3 in File 1; see Section “Analysis of sequencing data” in Materials and methods). We found a total of 200 SNMs and 136 small indels (<50 bp) across all end-point clones. The fact that we detected only a total of 336 SNMs/indels prevents us from examining more subtle effects as we did with LOH events. Thus, we restrict our analyses here to overall genome-wide effects.

We found that a typical C and D end-point clone accumulated 120/83 = 1.45 and 98/67 = 1.46 SNMs or indels, which is statistically indistinguishable from 118/79 = 1.49 mutations found in a typical W clone (Fig. 4, Table 1; $P = 0.843$ and $P = 0.629$ for C and D clones respectively; permutation test). Accounting for mutations that have been overwritten by LOH events (Sharp et al. 2018) introduces only a small correction (<3%) and does not change this result (Table 1; see Section “Analysis of sequencing data” in Materials and methods). Estimating the statistical power of our experiment to detect differences in mutation rates, we found that a detection of a 50% increase in mutation rate would have been almost certain, and a 30% increase would have been detected with probability 80% (Fig. 4a). Thus, we conclude that presence of an active CCGD or naked Cas9 increases point mutation rates in yeast (if at all) likely by less than 30% and almost certainly by less than 50%.

Discussion

In this study, we investigated how the presence of a CCGD element or the Cas9 gene without a dedicated gRNA in the genome affects the rates of LOH events and new mutations in yeast. We found no evidence that the overall LOH or mutation rates were elevated in strains carrying the CCGD or the Cas9 gene. Given the large size of our study, the absence of statistically significant effects on the overall LOH rate strongly suggests that these effects are small. Specifically, we estimate that the presence of a CCGD or the Cas9 protein alone increases the LOH and mutation rates likely by no more than 30%. By comparison, other studies found up to a 50-fold difference in the LOH event rates among yeast hybrid strains (Dutta et al. 2017; Overton and Kryazhimskiy 2025). Thus, the possibility of a 30% increase in the LOH event rate due to CCGD carriage is quite modest compared with that produced by natural variation. Similarly, an increase in the SNM rate of 30% is much smaller than the 5-fold variation in mutation rates among (nonmutator) yeast strains (Pankajam et al. 2020) and the 6-fold variation observed across the genome of a single yeast strain (Lang and Murray 2011).

While we find no detectable effects of CCGD or Cas9 on the genome-wide LOH rates, the presence of Cas9 possibly biases events to favor iLOH events and makes tLOH events shorter, though these results are only marginally significant. While we cannot rule out the possibility of Type I errors, this shift could point to a real biological effect. The presence of Cas9 may alter the probabilities with which the cell engages different DNA repair mechanisms, such that, for example, dsDNA breaks that occur closer to centromeres are more likely to result in iLOH events in C strains than in W strains.

Literature cited

- Adolfi A et al. 2020. Efficient population modification gene-drive rescue system in the malaria mosquito *Anopheles stephensi*. *Nat Commun.* 11:5553. <https://doi.org/10.1038/s41467-020-19426-0>.
- Akcakaya P et al. 2018. In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature.* 561:416–419. <https://doi.org/10.1038/s41586-018-0500-9>.
- Allen F et al. 2019. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol.* 37:64–72. <https://doi.org/10.1038/nbt.4317>.
- Arras SDM et al. 2022. Creeping yeast: a simple, cheap and robust protocol for the identification of mating type in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 22:foac017. <https://doi.org/10.1093/femsyr/foac017>.
- Baym M et al. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One.* 10:e0128036. <https://doi.org/10.1371/journal.pone.0128036>.
- Beaghton AK et al. 2019. Gene drive for population genetic control: non-functional resistance and parental effects. *Proc Biol Sci.* 286:20191586. <https://doi.org/10.1098/rspb.2019.1586>.
- Bier E. 2022. Gene drives gaining speed. *Nat Rev Genet.* 23:5–22. <https://doi.org/10.1038/s41576-021-00386-0>.
- Bloom JS et al. 2013. Finding the sources of missing heritability in a yeast cross. *Nature.* 494:234–237. <https://doi.org/10.1038/nature11867>.
- Boyle EA et al. 2021. Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes of target engagement. *Sci Adv.* 7:eabe5496. <https://doi.org/10.1126/sciadv.abe5496>.
- Champer J et al. 2017. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet.* 13:e1006796. <https://doi.org/10.1371/journal.pgen.1006796>.
- Champer J, Buchman A, Akbari OS. 2016. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nat Rev Genet.* 17:146–159. <https://doi.org/10.1038/nrg.2015.34>.
- Cradick TJ, Fine EJ, Antico CJ, Bao G. 2013. CRISPR/Cas9 systems targeting β -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 41:9584–9592. <https://doi.org/10.1093/nar/gkt714>.
- DePristo MA et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43:491–498. <https://doi.org/10.1038/ng.806>.
- DiCarlo JE et al. 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-cas systems. *Nucleic Acids Res.* 41:4336–4343. <https://doi.org/10.1093/nar/gkt135>.
- DiCarlo JE et al. 2015. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nat Biotechnol.* 33:1250–1255. <https://doi.org/10.1038/nbt.3412>.
- Dowd C. 2020. A new ECDF two-sample test statistic. *arXiv.* <http://arxiv.org/abs/2007.01360>
- Dowd C. 2023. Twosamples: fast permutation based two sample tests. <https://twosampletest.com/>.
- Dutta A et al. 2017. Genome dynamics of hybrid *Saccharomyces cerevisiae* during vegetative and meiotic divisions. *G3 (Bethesda).* 7: 3669–3679. <https://doi.org/10.1534/g3.117.1135>.
- Dutta A, Dutreux F, Schacherer J. 2021. Loss of heterozygosity results in rapid but variable genome homogenization across yeast genetic backgrounds. *eLife.* 10:e70339. <https://doi.org/10.7554/eLife.70339>.
- Esvelt KM, Smidler AL, Catteruccia F, Church GM. 2014. Concerning RNA-guided gene drives for the alteration of wild populations. *eLife.* 3:e03401. <https://doi.org/10.7554/eLife.03401>.
- Frankham R. 1995. Conservation genetics. *Annu Rev Genet.* 29: 305–327. <https://doi.org/10.1146/annurev.ge.29.120195.001513>.
- Fu Y et al. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.* 31: 822–826. <https://doi.org/10.1038/nbt.2623>.
- Gantz VM et al. 2015. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc Natl Acad Sci U S A.* 112:E6736–E6743. <https://doi.org/10.1073/pnas.1521077112>.
- Gantz VM, Bier E. 2015. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science.* 348:442–444. <https://doi.org/10.1126/science.aaa5945>.
- Garrood WT et al. 2021. Analysis of off-target effects in CRISPR-based gene drives in the human malaria mosquito. *Proc Natl Acad Sci U S A.* 118:e2004838117. <https://doi.org/10.1073/pnas.2004838117>.
- Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc.* 2:31–34. <https://doi.org/10.1038/nprot.2007.13>.
- Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu Rev Ecol Evol Syst.* 40:151–172. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173437>.
- Hoberecht L, Perampalam P, Lun A, Fortin J-P. 2022. A comprehensive Bioconductor ecosystem for the design of CRISPR guide RNAs across nucleases and technologies. *Nat Commun.* 13: 6568. <https://doi.org/10.1038/s41467-022-34320-7>.
- Hsu PD et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* 31:827–832. <https://doi.org/10.1038/nbt.2647>.
- Hum YF, Jinks-Robertson S. 2017. Mitotic gene conversion tracts associated with repair of a defined double-strand break in *Saccharomyces cerevisiae*. *bioRxiv.* <https://doi.org/10.1101/132167>
- James TY et al. 2019. Adaptation by loss of heterozygosity in *Saccharomyces cerevisiae* clones under divergent selection. *Genetics.* 213:665–683. <https://doi.org/10.1534/genetics.119.302411>.
- Kandul NP et al. 2021. A confinable home-and-rescue gene drive for population modification. *eLife.* 10:e65939. <https://doi.org/10.7554/eLife.65939>.
- Labib K, Hodgson B. 2007. Replication fork barriers: pausing for a break or stalling for time? *EMBO Rep.* 8:346–353. <https://doi.org/10.1038/sj.embor.7400940>.
- Lang GI, Murray AW. 2008. Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics.* 178:67–82. <https://doi.org/10.1534/genetics.107.071506>.
- Lang GI, Murray AW. 2011. Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol Evol.* 3:799–811. <https://doi.org/10.1093/gbe/evr054>.
- Li B et al. 2008. Heterologous expression of the TsVP gene improves the drought resistance of maize. *Plant Biotechnol J.* 6:146–159. <https://doi.org/10.1111/j.1467-7652.2007.00301.x>.
- Lin Y et al. 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* 42:7473–7485. <https://doi.org/10.1093/nar/gku402>.
- Loeillet S et al. 2020. Trajectory and uniqueness of mutational signatures in yeast mutators. *Proc Natl Acad Sci U S A.* 117: 24947–24956. <https://doi.org/10.1073/pnas.2011332117>.
- Makarova AV, Burgers PM. 2015. Eukaryotic DNA polymerase ζ . *DNA Repair (Amst).* 29:47–55. <https://doi.org/10.1016/j.dnarep.2015.02.012>.
- Malkova A et al. 2001. RAD51-independent break-induced replication to repair a broken chromosome depends on a distant enhancer

- site. *Genes Dev.* 15:1055–1060. <https://doi.org/10.1101/gad.875901>.
- Malkova A, Ivanov EL, Haber JE. 1996. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci U S A.* 93:7131–7136. <https://doi.org/10.1073/pnas.93.14.7131>.
- Mandegar MA, Otto SP. 2007. Mitotic recombination counteracts the benefits of genetic segregation. *Proc Biol Sci.* 274:1301–1307. <https://doi.org/10.1098/rspb.2007.0056>.
- Marshall JM, Akbari OS. 2018. Can CRISPR-based gene drive be confined in the wild? A question for molecular and population biology. *ACS Chem Biol.* 13:424–430. <https://doi.org/10.1021/acscchembio.7b00923>.
- Marshall JM, Buchman A, Sánchez C HM, Akbari OS. 2017. Overcoming evolved resistance to population-suppressing homing-based gene drives. *Sci Rep.* 7:3776. <https://doi.org/10.1038/s41598-017-02744-7>.
- McDonald MJ, Rice DP, Desai MM. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature.* 531:233–236. <https://doi.org/10.1038/nature17143>.
- McKenna A et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- Mehta A, Haber JE. 2014. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol.* 6:a016428. <https://doi.org/10.1101/cshperspect.a016428>.
- Newton MD et al. 2019. DNA stretching induces Cas9 off-target activity. *Nat Struct Mol Biol.* 26:185–192. <https://doi.org/10.1038/s41594-019-0188-z>.
- Nguyen Ba AN et al. 2022. Barcoded bulk QTL mapping reveals highly polygenic and epistatic architecture of complex traits in yeast. *eLife.* 11:e73983. <https://doi.org/10.7554/eLife.73983>.
- Oberhofer G, Ivy T, Hay BA. 2018. Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs. *Proc Natl Acad Sci U S A.* 115: E9343–E9352. <https://doi.org/10.1073/pnas.1805278115>.
- Overton MS, Kryazhimskiy S. 2025. Improved genotype inference reveals cis- and trans-driven variation in the loss-of-heterozygosity rates in yeast. *Genetics.* <https://doi.org/10.1093/genetics/iyaf274>.
- Pankajam AV et al. 2020. Loss of heterozygosity and base mutation rates vary among *Saccharomyces cerevisiae* hybrid strains. *G3 (Bethesda).* 10:3309–3319. <https://doi.org/10.1534/g3.120.401551>.
- Pattanayak V et al. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol.* 31:839–843. <https://doi.org/10.1038/nbt.2673>.
- Poplin R et al. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv.* <https://doi.org/10.1101/201178>
- Rode NO et al. 2019. Population management using gene drive: molecular design, models of spread dynamics and assessment of ecological risks. *Conserv Genet.* 20:671–690. <https://doi.org/10.1007/s10592-019-01165-5>.
- Roggenkamp E et al. 2018. Tuning CRISPR-Cas9 gene drives in *Saccharomyces cerevisiae*. *G3 (Bethesda).* 8:999–1018. <https://doi.org/10.1534/g3.117.300557>.
- Sharp NP, Sandell L, James CG, Otto SP. 2018. The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc Natl Acad Sci U S A.* 115:E5046–E5055. <https://doi.org/10.1073/pnas.1801040115>.
- Shyian M, Shore D. 2021. Approaching protein barriers: emerging mechanisms of replication pausing in eukaryotes. *Front Cell Dev Biol.* 9:672510. <https://doi.org/10.3389/fcell.2021.672510>.
- Smit AFA, Hubley R, Green P. 2023. RepeatMasker Open-4.0. 2013–2015. <https://repeatmasker.org/>.
- Smukowski Heil CS et al. 2017. Loss of heterozygosity drives adaptation in hybrid yeast. *Mol Biol Evol.* 34:1596–1612. <https://doi.org/10.1093/molbev/msx098>.
- Smukowski Heil C. 2023. Loss of heterozygosity and its importance in evolution. *J Mol Evol.* 91:369–377. <https://doi.org/10.1007/s00239-022-10088-8>.
- Sternberg SH et al. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature.* 507:62–67. <https://doi.org/10.1038/nature13011>.
- Sui Y et al. 2020. Genome-wide mapping of spontaneous genetic alterations in diploid yeast cells. *Proc Natl Acad Sci U S A.* 117: 28191–28200. <https://doi.org/10.1073/pnas.2018633117>.
- Symington LS, Rothstein R, Lisby M. 2014. Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. *Genetics.* 198: 795–835. <https://doi.org/10.1534/genetics.114.166140>.
- Tattini L et al. 2019. Accurate tracking of the mutational landscape of diploid hybrid genomes. *Mol Biol Evol.* 36:2861–2877. <https://doi.org/10.1093/molbev/msz177>.
- Tsai SQ et al. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol.* 33: 187–197. <https://doi.org/10.1038/nbt.3117>.
- Tsai SQ et al. 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods.* 14:607–614. <https://doi.org/10.1038/nmeth.4278>.
- Tutaj H, Pirog A, Tomala K, Korona R. 2022. Genome-scale patterns in the loss of heterozygosity incidence in *Saccharomyces cerevisiae*. *Genetics.* 221:iyac032. <https://doi.org/10.1093/genetics/iyac032>.
- Unckless RL, Clark AG, Messer PW. 2017. Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics.* 205:827–841. <https://doi.org/10.1534/genetics.116.197285>.
- van der Auwera G, O'Connor BD. 2020. Genomics in the cloud. O'Reilly Media.
- Vaserstein LN. 1969. Markov processes over denumerable products of spaces, describing large systems of automata. *Probl Peredachi Inf.* 5:64–72. <http://mi.mathnet.ru/ppi1811>.
- Vijayan N, Joshi S, Sarath P, Nishant KT. 2025. Loss of heterozygosity associated with ubiquitous environments in yeast. *PLoS Genet.* 21:e1011692. <https://doi.org/10.1371/journal.pgen.1011692>.
- Wang S, Jacobs-Lorena M. 2013. Genetic approaches to interfere with malaria transmission by vector mosquitoes. *Trends Biotechnol.* 31:185–193. <https://doi.org/10.1016/j.tibtech.2013.01.001>.
- Wernberg T et al. 2018. Genetic diversity and kelp forest vulnerability to climatic stress. *Sci Rep.* 8:1851. <https://doi.org/10.1038/s41598-018-20009-9>.
- Yin Y, Petes TD. 2013. Genome-wide high-resolution mapping of UV-induced mitotic recombination events in *Saccharomyces cerevisiae*. *PLoS Genet.* 9:e1003894. <https://doi.org/10.1371/journal.pgen.1003894>.
- Zheng DQ et al. 2016. Global analysis of genomic instability caused by DNA replication stress in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.* 113:E8114–E8121. <https://doi.org/10.1073/pnas.1618129113>.
- Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci U S A.* 111:E2310–E2318. <https://doi.org/10.1073/pnas.1323011111>.