# Outcrossing Complicates Mutation Purging by Trapping Single Nucleotide Polymorphisms in Structural Variant Mutations

R. Kapila[1,2*], S. Saber[1,2], R.K. Verma[1,2], G. Blanco[1,2], V.K. Eggers[1,2,] J.L. Fierst[1,2*]

[1]Department of Biological Sciences and [2]Biomolecular Sciences Institute

Florida International University, 11200 8th Street, Miami, 33199, FL, USA

*Corresponding author(s). E-mail(s): rkapila@fiu.edu; jfierst@fiu.edu

## Abstract

Outcrossing influences a population's ability to purge deleterious mutations, but its impact on larger structural genomic changes like insertions, deletions, and inversions remains unclear. To test this, we developed a theoretical model of structural variant (SV) mutation dynamics, implemented it in population simulations, and tested our predictions by exposing three *C. elegans* strains to ethyl methanesulfonate and formaldehyde to induce diverse mutations. Contrary to expectations for single nucleotide polymorphisms (SNPs), our model and whole-genome sequencing revealed that outcrossing could retain SVs in populations even after recovery from mutagenesis. The strain with highest levels of outcrossing accumulated the largest number of SVs, including megabase-scale rearrangements, and showed the highest transposable element activity. Moreover, SNPs persisted when embedded within retained SVs. These findings contradict classical population genetic expectations, showing that outcrossing can hinder SV removal and complicate mutation purging by trapping SNPs within SVs.

## Introduction

The accumulation of deleterious mutations can have severe consequences for populations, leading to long-term fitness decline or, in extreme cases, extinction[1]. The ability to effectively eliminate or purge harmful mutations is critical for reducing genetic load and ensuring long-term evolutionary fitness[2]. Based on their size and scale, mutations are broadly categorized into two types: single nucleotide polymorphisms (SNPs) and structural variants (SVs). SNPs are small-scale changes involving a single base pair, often resulting in localized effects on amino acids or regulatory elements[3]. In contrast, SVs are large-scale genomic changes including deletions, duplications, inversions, and translocations. These can impact larger portions of the genome, frequently altering multiple genes or regulatory networks[4].

While mutation purging operates on both SNPs and SVs, the process may be more straightforward for SNPs due to their smaller scale, localized effects, and simpler genomic dynamics[5]. In contrast, purging

SVs may be more complex due to their larger genomic footprint, reduced recombination, and pleiotropic effects[6]. Evolutionary theory has traditionally viewed the spectrum of genomic changes under the broad category of "mutation". Much of the research on mutation purging has disproportionately focused on SNPs as they are more easily and accurately detected[7]. However, the advent of large-scale genome sequencing has revealed the ubiquity of SVs. Despite the foundational role of mutation purging in evolutionary theory[8–10], our understanding of how it operates on SVs remains unexplored. Addressing this is essential to refining theoretical models, validating predictions with empirical evidence and understanding how mutation size and severity interact with established evolutionary predictions.

Mating systems influence mutation purging by affecting genetic diversity, recombination, and the expression patterns of deleterious mutations[11,12]. Current consensus in the field is that outcrossing populations, through recombination, create novel genetic combinations which facilitates the purging of deleterious variation[13]. However, outcrossing can also potentially dilute the effects of deleterious mutations and slow their purging[14]. Reduced recombination and genetic diversity in populations that reproduce through self-fertility may lead to the accumulation of mildly deleterious mutations through genetic drift[15]. Selfing however, can also increase homozygosity, exposing harmful recessive mutations to selection and enabling more effective purging[15]. These processes may differ for mutations of different physical sizes. SNPs, with largely localized effects, may be more directly exposed to selection. Purging deleterious SVs may be complicated by their larger size, impacts, and pleiotropic effects, each of which can vary significantly with mating system[16,17]. In this study we combine theoretical modelling with experimental evolution to address two key questions: First, how do quantitative levels of outcrossing influence the genome's ability to purge deleterious mutations? Second, how does this ability vary with mutation size and genomic location?

To investigate these questions, we developed a mathematical model describing the evolution of structural variants in the genome and conducted population simulations to predict how SV mutations behaved under varying levels of outcrossing. We experimentally tested our predictions using three strains of *Caenorhabditis elegans* which differ in natural frequency of males in the population: N2, AB1 and CB4856. As an androdiecious species *C. elegans* populations consist of both males and hermaphrodites, allowing for reproduction through either self-fertilization or outcrossing with males. While self-fertilization predominates in *C. elegans*, the frequency of males and outcrossing varies drastically across different strains[18] making *C. elegans* a prime system to test these ideas.

We subjected *C. elegans* strains to two contrasting mutagens, EMS to induce single nucleotide mutations and formaldehyde to introduce complex lesions in the genome[19]. We mutagenized worms for

five consecutive generations and allowed them to recover from induced mutations for three generations. We then tracked the population's relative fitness, changes in male frequency, and outcrossing rates to investigate their ability to cope with mutational input and purge deleterious mutations. We employed both long-read and short-read sequencing to quantify the size, location, and types of mutations that persist after recovery.

## Results

### *Theoretical Expectations*

Although there is a body of work modelling the influence of outcrossing on mutation in populations these models typically assume that mutations are occurring at the level of an individual SNP or broader, abstract 'allele' level[20,21]. Importantly, the assumptions of these models do not fit our current understanding of SVs in natural populations. Therefore, we developed a set of mathematical equations with assumptions better suited to what we currently know about SVs, describing evolution of SVs in the genome. The full model is described in the Methods; here, we outline critical assumptions.

Our model described a genomic system where mutations occur as SVs, either insertions, deletions or inversions. Each individual in a population of size K carried $\mathrm{n}$ structural mutations that occurred at a rate $\mu$ each generation and were removed at a rate $\nu$, for example by back mutation via gene conversion[22]. We did not explicitly model the physical size of mutations but assumed that mutations had discrete phenotypic effect sizes that cumulatively influenced fitness. The distribution of fitness effects was drawn from a gamma distribution, an empirically supported assumption[23–25]. With this function the majority of mutations had small effects on the phenotype and fitness and a few large effect mutations were possible (SFigure 1).

Individual SVs interacted to decrease fitness faster-than-linearly and the fitness landscape had a negative curvature or synergistic epistasis (SFigure 2). The fitness of an individual carrying $\mathrm{n}$ SVs was $\mathrm{w(n)} = \exp^{-(\mathrm{an}+(\mathrm{bn}^2)/2)}$ where $\mathrm{a}$ and $\mathrm{b}$ were parameters determining the curvature of the fitness landscape (Charlesworth 1990). We implemented this model in stochastic simulations using the SLiM 4.3 software[26] and measured the impact of outcrossing rates on SVs in the population.

Contrary to the predictions of population genetic models studying SNPs and alleles[20,21], our model predicted that outcrossing retains SVs in populations (Figure 1). Our fully self-fertile populations rapidly purged SVs while populations with >30% of individuals reproducing through outcrossing retained SVs throughout their evolution. Between these two extremes our model predicted that outcrossing could

have a quantitative impact on the retention of SVs. Increasing levels of outcrossing increased both the number of populations retaining SVs and the number of mutations retained.

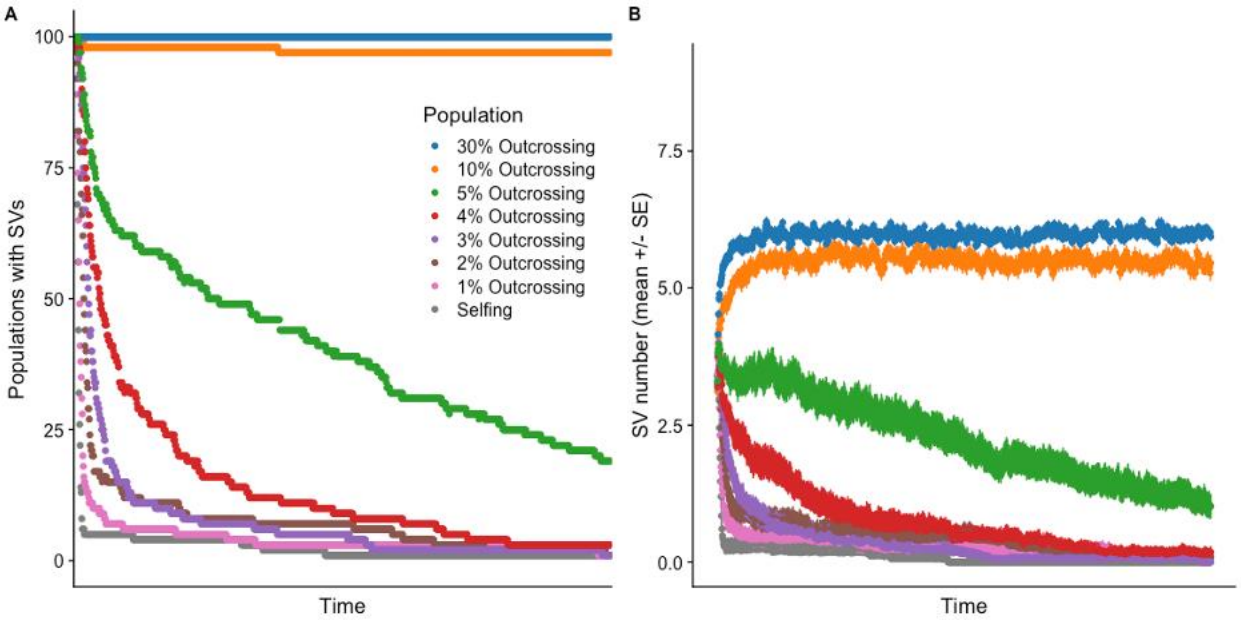



Figure 1. A) Simulated populations evolving with less than 10% of individuals reproducing through outcrossing do not retain SVs. B) The mean number of structural variant mutations is lower in populations with lower quantitative levels of reproduction through outcrossing. Here K=1000, a=0.001, b =0.0006, μ=0.01, ν=0.001 and recombination rate r=0.0001.

*Experimental Test of the Outcrossing Prediction*

To test the predictions of our theoretical model we subjected three laboratory strains of *C. elegans* (N2 (Bristol), AB1, and CB4856 (Hawaiian), obtained from the *Caenorhabditis* Genetics Center, University of Minnesota) to five generations of mutagenesis followed by three generations of recovery at standard laboratory plates with large and high-density populations. During both the mutagenesis and recovery phases, worms were maintained on standard laboratory plates. However, during the mutagenesis phase, worms were suspended in mutagen solutions for 4 hours every fourth day before being transferred back to standard plates. In contrast, during the recovery phase, worms were not exposed to the mutagen and were allowed to grow normally on standard laboratory plates. We measured both male frequency and outcrossing rates at two key time points: before and after mutagenesis. Additionally, we assessed relative fitness for each strain at three time points: before mutagenesis, directly after mutagenesis, and after recovery. Each strain was maintained in two distinct lineages: one mutagenized with EMS and the other with formaldehyde, with four replicate populations per lineage to ensure statistical rigor.

118 *Male frequency varied between C. elegans strains*

119 To assess the impact of five generations of mutagen exposure, we measured male frequency in both non-
120 mutagenized and 5 generation mutagenized *C. elegans*. Age-synchronized L4 hermaphrodites were
121 assayed individually for lifetime progeny sex ratio across replicate plates. We used a zero-inflated beta
122 regression model[27] to analyze male frequency, followed by pairwise post hoc comparisons using Tukey's
123 HSD comparison[28]. We found that non-mutagenized parents of CB4856 had the highest male frequency,
124 which was significantly greater than non-mutagenized parents of AB1 ($p < 0.0001$) and non-mutagenized
125 parents of N2 ($p < 0.01$).

126 When mutagenized with EMS for 5 consecutive generations, CB4856 showed a significant drop in
127 male frequency compared to non-mutagenized CB4856 parent ($p = 0.03$). When mutagenized with
128 formaldehyde for 5 generations, compared to parental populations, the drop in male frequency was not
129 significant ($p = 0.11$). For N2 non-mutagenized parental populations, male frequency was not significantly
130 different from N2 EMS ($p = 1.00$) or Formaldehyde treatment ($p = 1.00$). Similarly, AB1 Formaldehyde and
131 AB1 EMS populations did not have male frequencies different from the AB1 non-mutagenized parental
132 populations ($p = 1.00$ for both) or from any treatments of N2 (all $p$-values were $> 0.05$; Supplementary
133 Material Table 1; Fig. 2A).

134

135 *Outcrossing frequencies varied between C. elegans strains*

136 Outcrossing rates were assessed across strains by monitoring sex ratios in progeny from single L4 worm
137 pairs over four days. Plates exceeding strain-specific male frequency thresholds were classified as
138 outcrossed. CB4856 populations consistently exhibited the highest outcrossing frequencies under all
139 conditions, with peaks following mutagenesis (Fig. 2A). N2 populations, in contrast, displayed the lowest
140 outcrossing frequencies. AB1 populations were intermediate to these with notably higher values in EMS
141 mutagenized populations compared with both parent strains and the AB1 formaldehyde mutagenized
142 population. These findings suggest that mutagens differ in their ability to stimulate outcrossing behavior
143 (Fig. 2A). Notably this difference in the outcrossing rate likely led to a difference in effective population
144 sizes of these strains. The highest outcrossing rate of CB4856 likely led to the highest effective population
145 size both before and after mutagenesis among the three strains.

146 *Fitness decline varied between C. elegans strains and populations*

147 Relative fitness was assessed by pairing focal worms with a GFP-marked competitor and quantifying
148 progeny composition after 8 days. Fitness was compared across non-mutagenized, post-mutagenized, and
149 recovered lineages using automated image analysis (Supplementary Material Table 2). To validate the

accuracy of the image analysis pipeline, we randomly selected a subset of samples and manually counted individuals. This confirmed that the automated method produced reliable counts, with no observed false positives or false negatives. We found that three generations of recovery from mutagenesis were sufficient for each strain to recover pre-mutagenesis fitness levels for both the mutagens (Fig. 2B). For CB4856, relative fitness before mutagenesis was not significantly different from relative fitness after three generations of recovery for formaldehyde ($p$ = 1.00) and EMS treatments ($p$ = 0.18). Similarly, for AB1, post-recovery relative fitness was not significantly different from parental fitness for formaldehyde ($p$ = 0.94) or EMS ($p$ = 0.76). N2 populations also exhibited no difference between parental and post-recovery fitness for formaldehyde ($p$ = 1.00) or EMS (p = 1.00).

CB4856 fitness decreased slightly in EMS mutagenized populations but the effect was not significant ($p$ = 0.18). For formaldehyde mutagenized populations fitness levels were nearly identical before and after recovery ($p$ = 1.00). In contrast, this was not observed for N2 and AB1. For N2, pre-mutagenesis and post-recovery relative fitness was not different for formaldehyde ($p$ = 1.00) or EMS ($p$ = 1.00). Similarly, for AB1, pre-mutagenesis and post-recovery fitness under formaldehyde ($p$ = 0.94) and EMS ($p$ = 0.76) did not show changes in fitness.

N2 exhibited a distinct response compared to the other strains. Regardless of the mutagen used, N2 experienced a temporary increase in relative fitness immediately following mutagen exposure compared to its non-mutagenized parental levels (formaldehyde: $p$ = 0.03; EMS: $p$ = 0.46). However, this fitness gain was transient. During recovery, N2 population fitness returned to pre-mutagenesis levels (formaldehyde: $p$ = 1.00; EMS: $p$ = 1.00). Overall, three generations of recovery from mutagenesis were sufficient to restore fitness to pre-mutagenesis levels across all strains. However, CB4856 exhibited a persistent, albeit non-significant, reduction in fitness following EMS exposure, suggesting it may be less resilient to mutagenic stress compared to N2 and AB1.

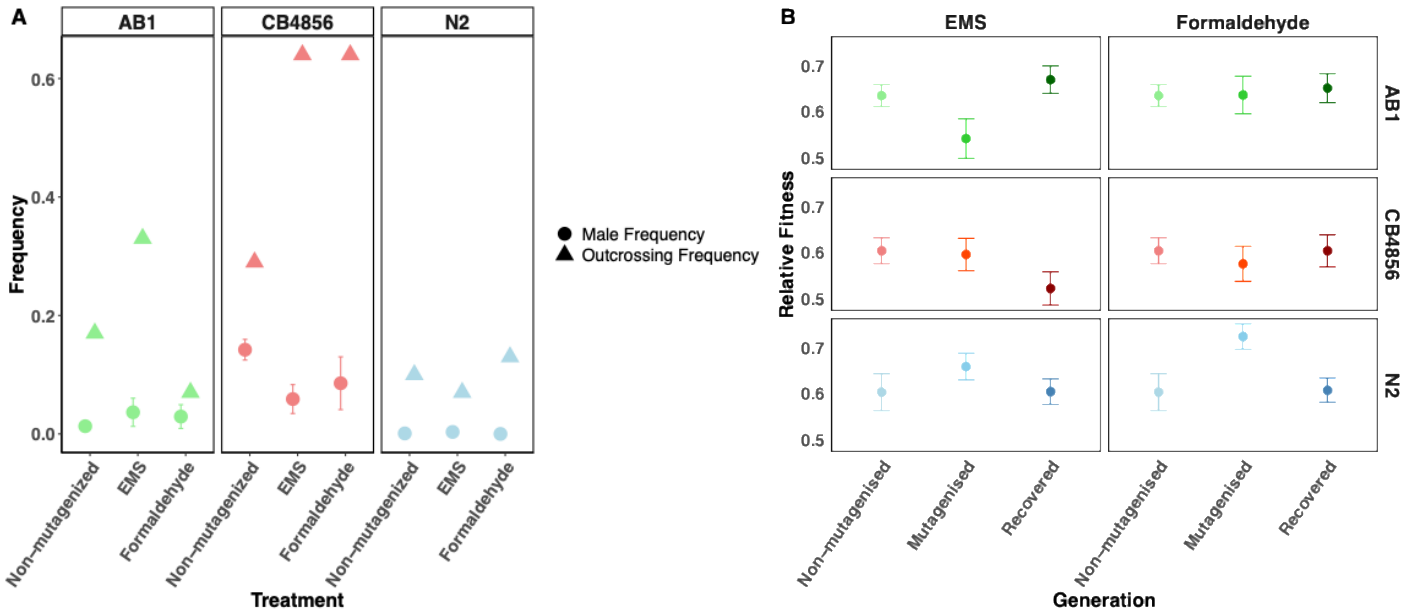


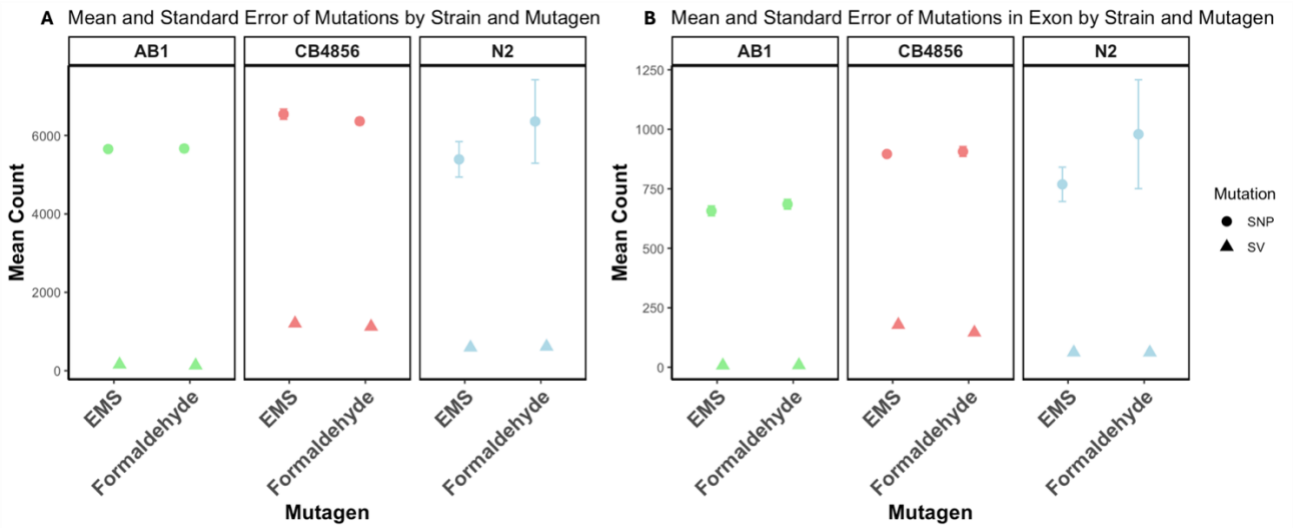

Figure 2. Male frequency, outcrossing behaviour, and relative fitness reveal strain-specific responses to mutagenic stress in C. elegans. (A) Mean male frequency across strains, represented by solid circles, before mutagenesis (non-mutagenized) and following EMS or formaldehyde exposure. CB4856 exhibited the highest male frequency across all treatments, followed by AB1 and N2. Outcrossing frequency, shown as solid triangles, was consistently higher in CB4856 across treatments, suggesting strain-specific differences in mating behaviour. (B) Relative fitness across strains and generations, comparing non-mutagenized (lightest shade), mutagenized (medium shade), and recovered (darkest shade) populations. AB1 (green) demonstrated the highest post-mutagenesis resilience, while CB4856 (red) showed the lowest recovery in fitness.

*Genomic analyses showed numerous SVs and SNPs were retained after recovery*

To comprehensively identify SV and SNP mutations we sequenced whole genome DNA libraries with both Pacific Biosystems long-read and Illumina short-read DNA sequencing. Short-read data was used for calling SNPs, whereas structural variants (SVs) were called from both short-reads and long-read data generating a high-confidence SV dataset capturing mutagen-induced genomic alterations.

We performed t-tests to compare de novo SNP counts between strains and found that CB4856 exhibited significantly higher susceptibility to mutagenesis compared to AB1 and comparable de novo SNPs to N2 when the mutagen was formaldehyde. Specifically, CB4856 showed a 1.12-fold increase in de novo SNPs compared to AB1 ($t=5.55$, $p=0.0014$) and comparable SNPs to N2 ($t < 0.01$, $p= 0.996$) under formaldehyde treatment. In contrast, with EMS as the mutagen, CB4856 showed a significantly higher (1.16-fold) de novo SNP count when compared with AB1 ($t = 6.79$, $p < 0.01$) and a marginally higher (1.21-fold) de novo SNP count when compared with N2 ($t = 2.45$, $p = 0.05$) as identified through short-read sequencing (Fig. 3A).

198 Due to marked heteroscedasticity observed in SV counts we performed Mann–Whitney U tests
199 (Wilcoxon rank-sum tests in R) to compare de novo SV counts among strains. Under EMS treatment,
200 CB4856 populations had significantly higher de novo SV counts compared with AB1 populations (W=16, *p*
201 = 0.03) and compared to N2 (W=16, *p*=0.03). Similarly, with formaldehyde treatment, CB4856 showed
202 significantly higher de novo SV counts compared to AB1 (W=16, *p*=0.03) and compared to N2 (W=16,
203 *p*=0.03) (Fig. 3A).



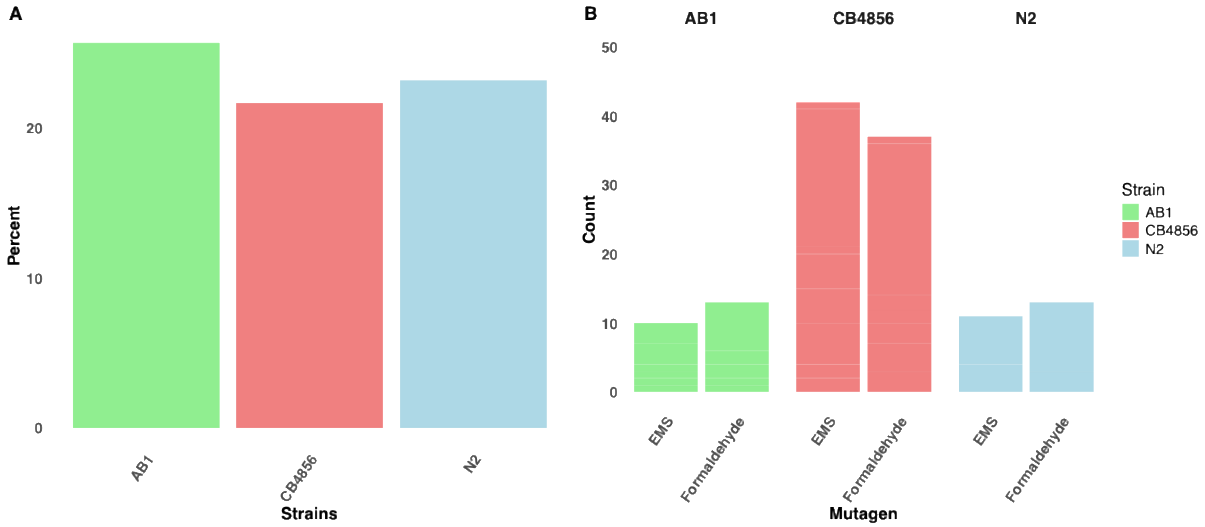

204
205 Figure 3: Comparison of de novo SNPs and SVs Mutations and Their Localization in Exons Across Strains
206 (A) Mean counts of de novo SNPs (Circles) and SVs (Triangles) induced by EMS and Formaldehyde across
207 AB1 (Green), CB4856 (Red), and N2 (Blue). De novo SNPs are significantly more abundant than de novo
208 SVs under both mutagen treatments. CB4856 has the maximum number of de novo SNPs and de novo
209 SVs. (B) Mean number of de novo SNPs and de novo SVs localized in exons for the same strains and
210 mutagen treatments, with de novo SNPs again being more prevalent than de novo SVs and CB4856 having
211 more SVs in exons than the other two strains and more SNPs in exons than AB1. Compared to N2, CB4856
212 had more SNPs in exon than N2 when treated with EMS and comparable SNPs in exon when treated with
213 Formaldehyde.
214

215 *CB4856, the strain with highest male frequency and outcrossing rates, harbored high numbers of both*
216 *SNPs and SVs in exons*

217 The higher number of de novo SNPs in CB4856 was also seen in the increased number of SNPs that were
218 present in exons. When formaldehyde was used as the mutagen, CB4856 had 2.20 times more SNPs in
219 exons than AB1 (W=0, *p*= 0.03) and 1.12 times more than N2 (W=12, *p*=0.34). With EMS as the mutagen,
220 CB4856 had 2.26 times more SNPs in exons compared to AB1 (W=0, *p*=0.03) and 1.38 times more SNPs in
221 exons compared to N2 (W=16, *p*=0.03) (Fig. 3B).

222 Similar trends were observed with exons affected by SVs. AB1 showed the most resilience with
223 the least number of exons affected by the SVs, whereas CB4856 had the maximum number of SVs affecting

224 exons. When formaldehyde was used as mutagen, AB1 had 16.25-fold fewer SVs in exons than CB4856
225 (W = 0, p = 0.02) and 6.92-fold fewer SVs than N2 (W = 16, p = 0.03). With EMS as the mutagen, CB4856
226 had 22.31-fold more SVs affecting the exons compared to AB1 (W = 0, p = 0.02), and N2 had 7.81-fold
227 more exons affected by SVs than AB1 (W = 16, p = 0.03) (Fig. 3.B).
228
229 *Active transposable elements were responsible for a small portion of SV mutations*
230 SV mutations can be associated with transposable element (TE) activity. To test this, we used
231 TransposonUltimate[29] to annotate repeats and transposable elements in our whole genome sequences.
232 We compared repeats and TEs present in our populations with SVs before and after mutagenesis to focus
233 on de novo TE-associated SV mutations and potentially active TEs[29]. The total percentage of the genome
234 annotated as repetitive was 25.71% in AB1, higher than that identified in N2 (23.19%) and CB4856
235 (21.68%). Despite this, the total number of transposition events was highest in CB4856 (Fig. 4A). CB4856
236 exhibited 2.85-fold more TE events than N2 and AB1 when treated with formaldehyde, and 3.82-fold more
237 TE events than N2 and 4.20-fold more TE events than AB1 when treated with EMS (Fig. 4B). While some
238 TEs were strain-specific, some like the DNA transposon Zator (Fig. 5B) were consistently mutation-
239 associated in all populations and across both mutagens, indicating that Zator may be commonly activated
240 by mutagenic stress in *C. elegans* (Fig. 5B). Interestingly, some TEs, such as Copia, Gypsy, and Sola, were
241 uniquely active in CB4856 and nearly absent from mutation characterization in N2 and AB1 populations.



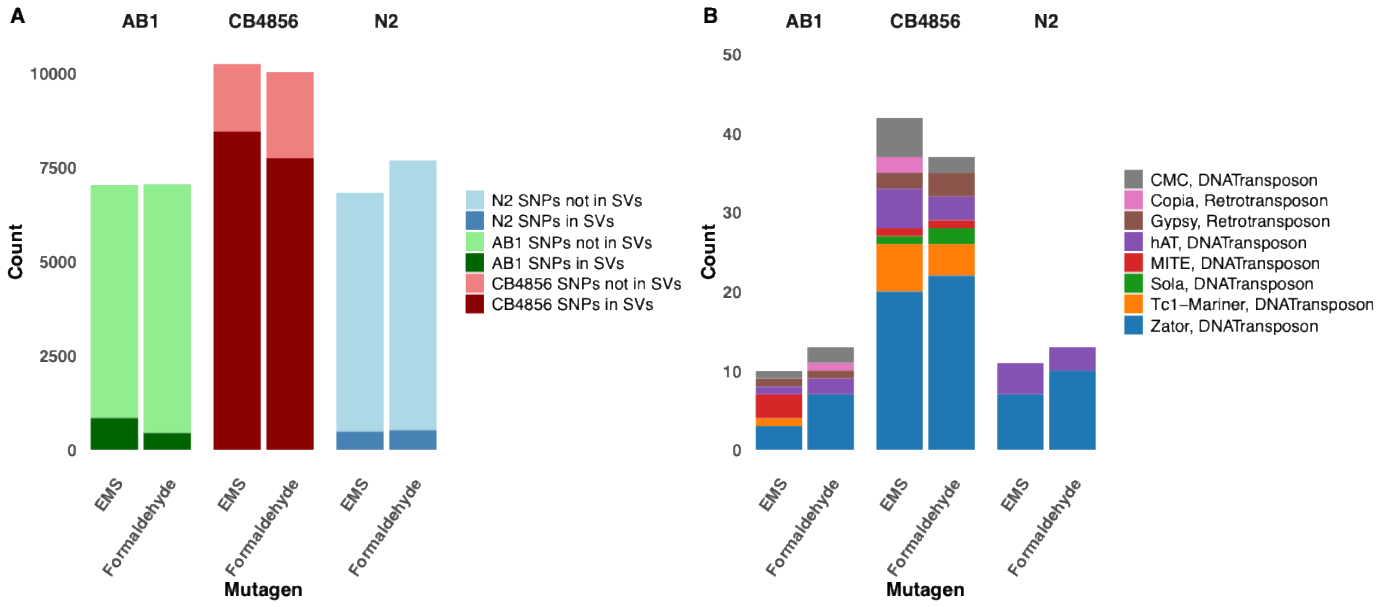

242
243 Figure 4: Transposable Element (TE) Activity and Genomic Repeat Content Across Strains and Treatments.
244 **(A)** Percentage of the genome annotated as repetitive elements across the three C. elegans strains (AB1,
245 CB4856, and N2). **(B)** Total TE transposition events observed post-mutagenesis, showing higher activity in
246 CB4856 compared to AB1 and N2 under both EMS and formaldehyde treatments.
247

248 *SNP Enrichment Within SVs Varies by Strain and Mutagen*

249 CB4856 had the highest number of SNPs trapped within SVs (77.13% for formaldehyde and 82.46% for
250 EMS treatment; Fig. 5A). This strain, characterized by high male frequency and outcrossing (Fig. 2A, B),
251 also had a high intersection of these mutations with exonic regions (Fig. 3B). In contrast, while the total
252 number of SNPs in AB1 and N2 strains appears similar, a closer examination reveals that AB1 harbors a
253 higher proportion of its SNPs within structural variants (SVs) compared to N2 when the mutagen was EMS
254 and almost comparable numbers when the mutagen was formaldehyde. (6.36% vs. 6.80% for
255 formaldehyde and 12.01% vs. 7.12% for EMS treatment). While most SVs were small (<$10^4$ bp), CB4856
256 harbored a notable proportion of larger SVs compared to N2 and AB1 under both EMS and formaldehyde
257 treatments, suggesting elevated genome instability in this strain (Fig. 6).

258

259




260

261 Figure 5: Mutational and TE landscape across strains and mutagens **(A)**: Number of SNPs, SVs and SNPs
262 in SVs for AB1, CB4856 and N2; **(B)**: Classification and count of active TE families contributing to
263 transposition events across strain-treatment combinations. DNA transposons such as Zator and Tc1-
264 Mariner were among the most active elements, with CB4856 showing the broadest TE activity profile.

265

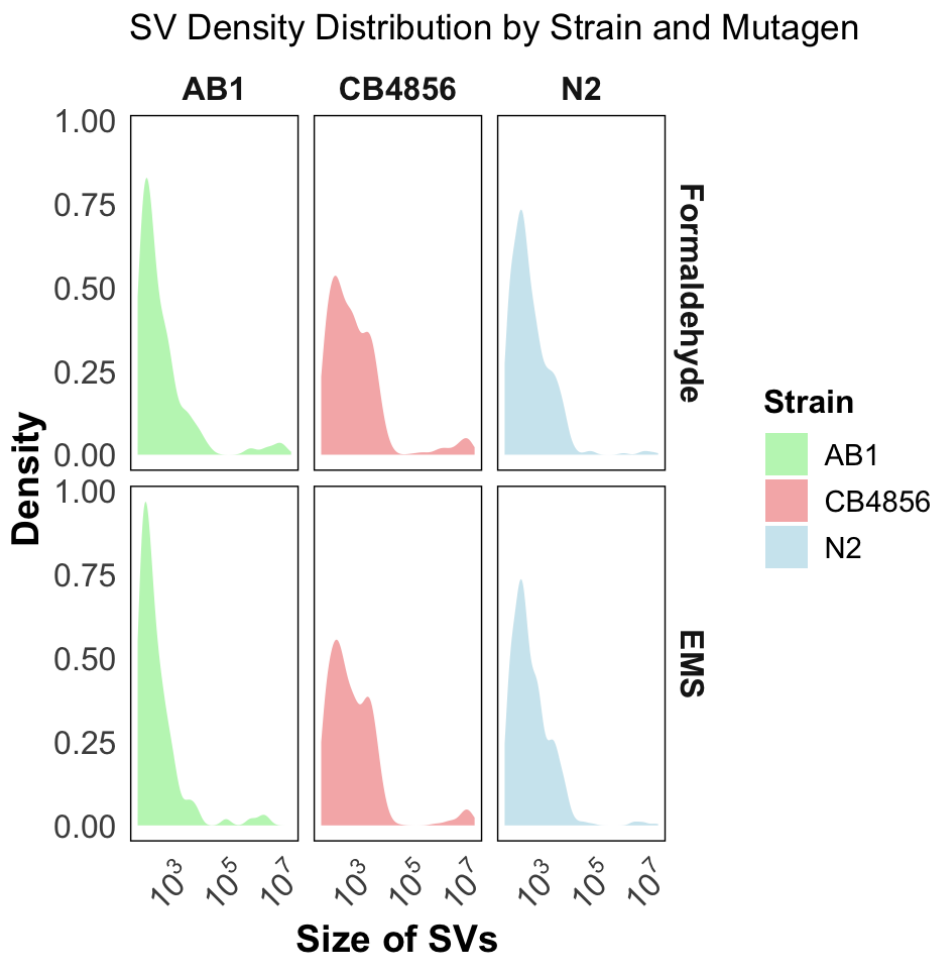



Figure 6: Size distribution of structural variants (SVs) across C. elegans strains and mutagens. Density distribution of SVs sizes for all the three strains for both the mutagens. x- axis Size of SVs, Y-axis: density of SVs.

*Genome-wide structural variants and transposon activity reveal distinct patterns*

The distribution of structural variations (SVs) was almost uniform across all chromosomes, regardless of strain or mutagen used (Fig. 7). In the N2 formaldehyde-treated population, insertion mutations created 471,238 nucleotides of new genetic material, approximately 0.48% of the 100.8 Mb C. elegans genome (Supplementary Material Table 3). In comparison, 74,044 nucleotides (0.06%) were deleted, 44,038 nucleotides (0.04%) were duplicated, and 1,289,368 nucleotides (1.29%) were affected by inversions. Mutations in the EMS-mutagenized N2 populations were of similar scale.

AB1 populations showed the smallest number of nucleotides affected by SV mutations. In formaldehyde-treated AB1, 35,468 nucleotides were inserted and 46,363 deleted. The notable exception was inversion mutations, which spanned 2,539,375 nucleotides in formaldehyde-treated and 4,181,603

nucleotides in EMS-treated populations. CB4856 populations harbored substantially more DNA affected by structural variation. Roughly 1 Mb of new DNA was inserted under both treatments. Compared to AB1, CB4856 had 14.11-fold more insertions under formaldehyde and 15.18-fold more under EMS. Relative to N2, it had 1.50-fold and 1.66-fold more insertions under formaldehyde and EMS, respectively. CB4856 also had large amounts of duplicated DNA: 1.2 Mb under formaldehyde and 2.1 Mb under EMS. Strikingly, inversion mutations accounted for 49.7 Mb—nearly half the genome—in formaldehyde-treated CB4856, including large 13 Mb and 8 Mb inversions, with additional megabase-scale and smaller rearrangements. In EMS-treated CB4856, inversions spanned 32.8 Mb—almost a third of the genome (Fig. 7.A).

Given the potential of transposable elements to contribute to structural variation, we further evaluated TE-associated mutations across strains and treatments. TE-associated duplications and inversions were less frequent than deletions across all populations (Fig. 7.B). Many TE-associated mutations occurred on chromosome IV, particularly in the CB4856 and N2 populations (Fig. 7.B). Additionally, we found shared transposition events between the two mutagens. The highest overlap between formaldehyde and EMS TE-associated mutations was observed in N2 where 33.3% of the mutations were shared. In CB4856 17.9% of the TE-associated mutations were shared between EMS and formaldehyde treated populations while in AB1 it was just 4.5%. These findings indicate that while transposable elements contribute to structural variation, they do not account for the full spectrum of observed SVs, and their activity patterns are distinct in both scale and genomic distribution.

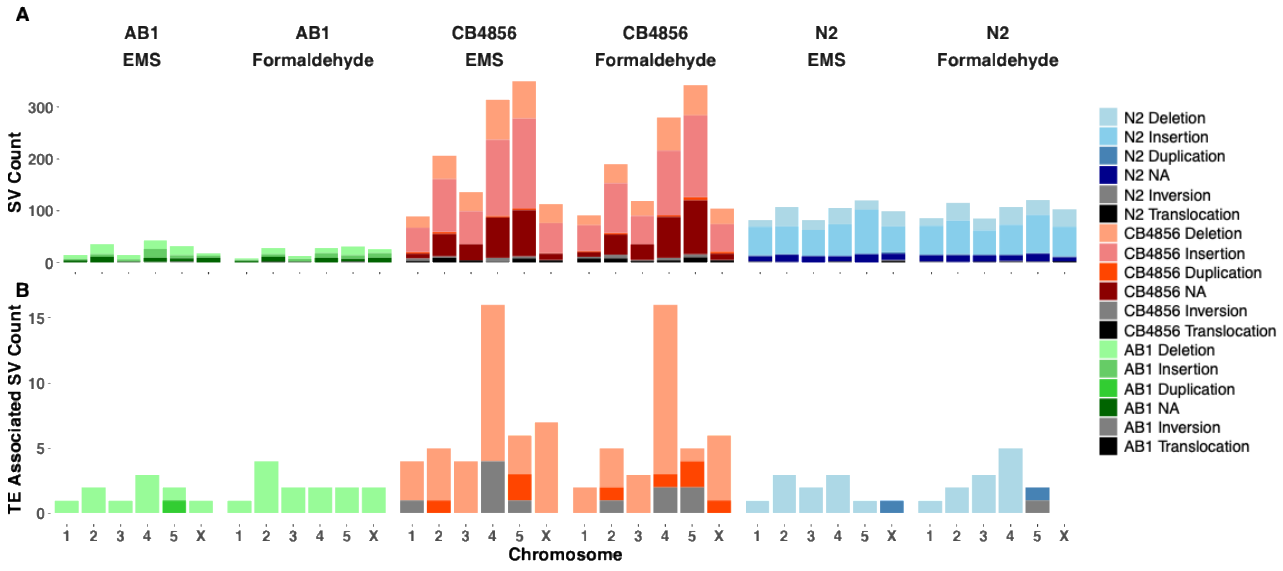



Figure 7: Chromosomal distribution of SVs and TE-associated SVs across strains and mutagens. (A) Stacked bar plots showing the total count and types of SVs including Deletions, Insertions, Duplications, Inversions,

Translocations, and Unclassified (NA) across chromosomes 1–5 and X, for each strain (AB1, CB4856, N2) and mutagen (EMS and formaldehyde). (B) TE associated SVs were primarily Deletions.

## Discussion

In this study we investigated how quantitative levels of outcrossing influence the genome's ability to purge deleterious mutations and how this ability varies with mutation size and genomic location. By developing a theoretical model to explore the evolution of structural variants (SVs) in populations with varying outcrossing rates and leveraging the unique mating system of *C. elegans*, we were able to create testable predictions and directly answer our questions with quantitative analyses. CB4856, characterized by highest male frequency, outcrossing rate and elevated TE activity, showed a trend toward reduced fitness after mutagenesis and recovery. While this decline was not statistically significant, it contrasted with the complete recovery seen in AB1 and N2, potentially reflecting strain-specific differences in genomic resilience. Mutation analysis agreed with our theoretical predictions and showed that CB4856 populations retained the highest number of SVs. Importantly, we found many SNPs were trapped within SVs and overlapped with exonic regions, making it difficult to purge these variants through recombination and outcrossing. Overall, these findings support the predictions of our theoretical model and show that higher outcrossing rates may in fact hinder the purging of larger genomic changes such as TEs and SVs.

*C. elegans* is one of the most tractable model systems and has emerged as a powerful model organism for unraveling the intricate dynamics of mutation accumulation. Traditionally, researchers have employed two primary approaches to investigate this phenomenon. The first approach involves manipulating DNA repair pathways, creating "mutation accumulation" lines with compromised repair mechanisms [30]. This allows for controlled observation of how increased mutation rates impact population fitness and adaptation over generations. The second approach takes a more observational stance, tracking the accumulation and consequences of naturally occurring mutations over extended periods[31]. Both approaches have yielded valuable insights into the remarkable ability of *C. elegans* populations to tolerate and potentially recover from mutational stress. However, despite these advancements, a crucial aspect remains underexplored: how mating system variation, reproductive strategies, and underlying genetic architecture interact with mutation size and its impact on genome-wide fitness.

We explored these factors in our theoretical model and predicted that increased levels of outcrossing would retain SVs. The number of deleterious mutations segregating in our simulated outcrossing populations was small, ranging between 10 and 20 mutations per individual. Importantly, the phenotypic effect size of the segregating mutations was also small. In a fully outcrossing simulated population, there was an average of 15 mutations per individual after 100,000 generations, each with an

333 average phenotypic effect size of 0.0035. Larger mutation effect sizes were rapidly purged from
334 outcrossing populations. However, deleterious mutations with small effects on the phenotype and fitness
335 were stably maintained in the simulated outcrossing populations for long periods.

336 We found that similar to our theoretical predictions, the *C. elegans* strain with the highest levels
337 of outcrossing, CB4856, retained the highest number of SVs. Our *C. elegans* strains displayed small
338 depressions in relative fitness, indicating that these genomic mutations also had small deleterious effects.
339 The mechanisms by which genomes cope with mutational threats are multifaceted and depend heavily
340 on the type of lesion incurred. Point mutations like SNPs predominantly affect gene expression and
341 protein structure and are typically repaired through efficient pathways like base excision repair (BER) or
342 nucleotide excision repair (NER). In contrast, complex lesions like SVs (structural variants) pose a more
343 significant challenge to genome stability, often leading to chromosomal aberrations[32]. Repairing these
344 intricate injuries often relies on more intricate mechanisms like homologous recombination (HR) and non-
345 homologous end joining (NHEJ)[33]. In our study we used two types of mutagens that are known to affect
346 the genome in different ways. EMS is known to cause more point mutations and formaldehyde is known
347 to cause complex lesions. We expected that EMS should be creating more SNPs and formaldehyde should
348 be causing more structural variations in the genome. However, we found that long term mutagenic abuse
349 (5 consecutive generations in our case) with either of the mutagen had similar mutational effects on the
350 *C. elegans* genome and resulted in SNPs and SVs ranging from 10s of nucleotides to multiple megabases
351 in length.

352 The potential impact of male frequency and outcrossing on mutational burden remains highly
353 debated. Studies in dioecious species suggest that intense male competition for mates drives stronger
354 selection against deleterious mutations, potentially benefiting populations facing genomic stress[34].
355 However, bottlenecks arising from such strong selection can ironically counteract this advantage by
356 allowing the accumulation of detrimental mutations. Additionally, the relevance of these findings to
357 androdiecious species like *C. elegans* is unclear. Due to their self-fertilizing nature, *C. elegans* males might
358 not experience the same level of selective pressure as those in dioecious species[35]. Consequently,
359 increased male frequency in *C. elegans* might not directly translate to mutational resilience. Nevertheless,
360 higher male proportions could offer another advantage: increased outcrossing, which can introduce new
361 genetic variation. This influx of novel alleles could enhance the population's adaptability and resilience in
362 the face of unpredictable challenges like ongoing mutagenic stress. Moreover, mutation purging
363 processes differ significantly between self-fertilizing and outcrossing mating systems, particularly for SNPs
364 and SVs. In self-fertilizing populations, increased homozygosity facilitates rapid purging of highly

365 deleterious mutations from the genome. However, reduced genetic variation weakens selection against
366 slightly deleterious mutations, potentially leading to the accumulation of slightly deleterious SNPs and
367 SVs. In contrast, outcrossing populations can maintain both highly deleterious and slightly deleterious
368 mutations if they are recessive.

369 Our results suggest that there are strain- and mutagen-specific effects, with CB4856 showing the
370 highest male frequency among the three strains. However, it exhibited the worst relative fitness after
371 three generations of recovery from mutagenic treatment and carried the highest mutational load for both
372 mutagens. AB1, in contrast, had second largest male frequency but was least affected by mutagenic abuse
373 and had the minimum mutagenic load in the genome when compared with the other two strains. N2 had
374 the lowest male frequency but relative fitness during mutagenesis, which dropped in recovery. These
375 results suggest there might be some costs associated with the mechanisms that resulted in N2s high
376 relative fitness in an unstable environment.

377 Strain-specific differences in coping with mutagenic stress may reflect distinct genomic
378 architectures. For example, in AB1 few SVs had high proportions of SNPs trapped within them and
379 overlapped infrequently with exonic regions. This genomic configuration, where SNPs are more frequently
380 associated with SVs but largely avoid exonic regions, could enhance a population's genomic resilience by
381 minimizing disruptions to exonic regions. This "buffering effect" could facilitate resilience in response to
382 mutagenic stress. SV-associated regions could serve as hotspots for repair or regulation, further
383 contributing to fitness and survival under stress.

384 CB4856 had the highest number of de novo SNPs and SVs. A high proportion of SNPs were trapped
385 within SVs and a significant number of these mutations intersected with exonic regions, potentially
386 disrupting essential genes. This genomic configuration likely explains CB4856's reduced resilience under
387 mutagenic stress, as the interplay between structural and single nucleotide variants can amplify the
388 deleterious effects of mutations, complicating the purging process. The broad and evenly distributed
389 mutational landscape especially for SV ranging from 1,000 base pairs to $10^7$ base pairs (Fig. 6) may favor
390 greater genomic plasticity, improving a species generation of genetic variation and potential adaptation
391 in uncertain environments. However, CB4856 experienced the most pronounced decline in fitness
392 following mutagenesis. The delicate balance between evolutionary plasticity and mutational resilience
393 may tip toward instability under the stress of chemical mutagenesis.

394 N2 populations showed relatively stable relative fitness responses without significant increases in
395 male frequencies or outcrossing frequencies and relatively low genomic responses to mutagenesis. This
396 suggests that N2 relied heavily on DNA repair mechanisms, rather than increased genetic variation

397 through outcrossing, to mitigate the negative effects of induced mutations. The combination of self-
398 fertilization and the retention of small SVs may minimize genomic disruptions, ensuring genomic integrity
399 over generations. This may balance minimizing risk through self-fertilization and small SV accumulation,
400 while still allowing for some level of genomic flexibility. This type of response may allow an organism to
401 persist in environments where maintaining genomic integrity is essential for survival. Interestingly, the
402 relative fitness of N2 with both the mutagens increased which came back to the pre-mutagenizes levels
403 after recovery phase. This suggest that the initial increase in fitness might come with a long-term cost,
404 potentially limiting its adaptive potential over time.

405 Transposable elements (TEs) can play a critical role in shaping genomic resilience by introducing
406 genetic variation and driving structural changes in response to stress[36]. Increased outcrossing has been
407 associated with elevated TE activity, likely due to the enhanced recombination rates and reduced
408 efficiency of silencing mechanisms in outcrossing populations[37,38]. Our results highlight significant strain-
409 specific differences in TE activity post-mutagenesis. CB4856, with the highest outcrossing frequency,
410 exhibited the highest transposition rates, reflecting its propensity for increased genomic instability under
411 mutagenic stress. Interestingly, although AB1 had the highest number of annotated transposons, its
412 transposition rates were lower, suggesting fewer active TEs. Elevated TE activity in CB4856, coupled with
413 unique transposons such as Copia, Gypsy, and Sola, may indicate a potential for adaptability through
414 increased genetic variation and a capacity to respond to environmental challenges. Interestingly, N2
415 exhibited the highest overlap in TE events between mutagens (33.3%), suggesting common TE activations
416 under stress. Across all strains, chromosome IV emerged as a TE hotspot, likely reflecting shared genomic
417 characteristics that tolerate or facilitate transposition in this region.

418 Increased outcrossing enhances recombination rates, creating genomic instability that can
419 activate TE mobilization[39]. Additionally, the introduction of genetic variation through outcrossing may
420 temporarily disrupt TE silencing mechanisms, further contributing to higher TE activity in CB4856[40]. These
421 results indicate that while higher outcrossing may promote the generation of genetic diversity, it can also
422 trigger processes that increase susceptibility to mutagenic stress. This susceptibility may stem from strain-
423 specific genomic architectures, the accumulation of deleterious mutations, or the difficulty of purging
424 large-scale genomic changes under conditions of frequent outcrossing.

425 In conclusion, we found strain specific fitness and mutation responses which may reflect larger
426 patterns of evolutionary response. Although the N2 "Bristol" strain originally collected by Sydney Brenner
427 (Brenner 1974) has been developed as a model system its fitness, behavior and genomics differ
428 substantially from other strains of *C. elegans*. N2 is a highly lab adapted strain with little genetic variation,

429 few males and low outcrossing frequencies[41]. In contrast, CB4856 and AB1 are more recently collected
430 "wild worms" with substantial genomic differences[42] and high natural variation in male frequency[43] and
431 outcrossing [44]. The spectrum of natural variation in *Caenorhabditis* will continue to be an important
432 resource in scientific understanding of mutation.
433
434 **Acknowledgements**
435 The authors would like to thank the members of the Fierst lab and Jessica Gonzalez, Jason Pienaar and
436 Jesualdo Fuentes-Gonzalez for suggestions and careful feedback.
437
438 **Funding**
439 This work was supported by NSF award 2225796 and NIGMS award R35GM147245 to J.L.F.
440
441 **Data Availability**
442 Scripts, bioinformatic workflows and analyses are available at github.com/RohitKapila. Python code used
443 for image analysis is available at github.com/FierstLab/WormImage Fitness data are available as
444 Supplemental Materials associated with the manuscript. Sequence reads are available at the NCBI SRA
445 under BioProject PRJNA1201181.
446
447 **Author contributions**
448 J.L.F. secured funding. R.K. and J.L.F. designed the experiments. R.K., with assistance from S.S., performed
449 the fitness assays, and, with assistance from G.B., conducted the male frequency and outcrossing rate
450 assays. V.K.E. carried out the transposable element analysis. R.K. performed the genomic analysis with
451 guidance from J.L.F. R.K.V. developed the Python code for worm counting. J.L.F. developed the theoretical
452 model and conducted the population simulations. R.K. and J.F.K. co-wrote the first draft. R.K. and J.F.K.
453 prepared the final manuscript with contributions from V.K.E.
454
455 **Methods**
456
457 *Theoretical model*
458 In order to model the evolution of structural variant mutations in a population we adapted a model of
459 transposable element mutations developed by Dolgin and Charlesworth (2006, 2008). This framework
460 built on previous work[45,46] studying the population dynamics of transposable elements. The fitness of an

individual was determined by the number of mutations they carried and calculated according to a synergistic fitness function where fitness declined faster-than-linearly as mutations accumulated in the genome. This synergistic fitness function was a fundamental assumption in theoretical work addressing the evolution of sex and mutation-selection balance [46,47]. Importantly, free recombination is assumed between each mutation.

We adapted this model in several ways. We envisioned a genomic system where structural variant mutations could take the form of insertions, deletions or inversions and, as described in the Results above, each mutation carried a discrete phenotypic effect size drawn from a gamma distribution. The phenotypic effects of individual mutations cumulatively determined fitness. Following Dolgin and Charlesworth (2006) the change in the population mean mutation number was $\Delta\bar{n} \approx V_n\ (\partial \ln w_\bar{n})/(\partial \bar{n}) + \bar{n}\ (\mu - \nu)$. Here, $V_n$ was the variance in copy number in the population and $w_\bar{n}$ was the fitness of an individual with $\bar{n}$ SV mutations in their genome.

Mutational effect sizes were $\Gamma(k, \theta)$ random variables giving $\Delta n \approx n\ (\theta\ (\partial \ln w_n)/(\partial n) + \mu - \nu)$. At equilibrium (defined as $\Delta n = 0$) the mean number of mutations carried in an individual's genome was $n = (\mu - \nu - \theta\, a)/(\theta\, b)$ where $a$ and $b$ were the parameters determining the curvature of the fitness landscape (Charlesworth 1990). According to this, we expected that the number of SV mutations segregating in a population at equilibrium would be inversely proportional to the variance of mutational effect sizes. With $\theta = 2, \mu = 0.01, \nu = 0.001, a = 0.001$ and $b = 0.0006$ we expected a mean of 5.83 SV mutations in a fully outcrossing population.

*Population simulations*

To quantitatively examine the influence of outcrossing on SV purging and retention we implemented our theoretical model in population simulations using the software SLiM 4.3[48]. We constructed dioecious populations containing equal numbers of males and females and self-fertile populations with a proportion of individuals reproducing through outcrossing and a proportion reproducing through selfing. Each population started at a specified size $K$ and evolved through overlapping generations with reproduction and fitness differences creating variable population sizes. Each individual had a diploid chromosome 10,000 nucleotides in length. Individuals were randomly chosen for reproduction according to a uniform function. Recombination occurred along the chromosomes with uniform probability at a specified rate r. After reproduction SV mutations occurred at a rate $\mu$ and were removed at a rate $\nu$. The probability of a SV mutation occurring at any location was uniform across the chromosome and each SV mutation conferred a small deleterious phenotypic effect drawn from $\Gamma(k, \theta)$.

Our simulated populations evolved for 100,000 generations with the parameters specified in our deterministic equations, $k = 1, \theta = 2, \mu = 0.01, \nu = 0.001, a = 0.001$ and $b = 0.0006$ We evolved populations under a range of evolutionary scenarios including recombination rates $r$=0.5-1x10$^{-5}$, $N$=100 to 5,000 and levels of self-fertility ranging from 0 to 100%.

*Strains and mutagens*

Three C. elegans strains with varying male frequencies were chosen: N2 (Bristol), CB4856 (Hawaiian), and AB1. ST2 served as a common competitor in relative fitness assays. All strains were obtained from the Caenorhabditis Genetics Center and maintained frozen until the experiment. Frozen strains were thawed on 75 mm × 13 mm NGM plates seeded with E. coli OP50. After one generation (~3.5 days at 20 °C) for recovery, each strain was divided into nine replicate populations. Prior to the experiment, replicates were synchronized by bleaching. Over five generations, four replicates per strain were mutagenized with formaldehyde, four with EMS, and one remained untreated as a non-mutagenized control. All replicates for a given strain were handled in parallel on the same day using the same batch of plates to minimize intra-strain variability due to plate effects. Following the mutagenesis phase, all populations were frozen and stored at −80 °C for downstream analyses.

*Mutagenesis*

Four replicates per strain were treated with 1mM Ethyl methane sulfonate (EMS) in M9 solution, while the other four replicates were treated with 0.1% (w/v) Formaldehyde solution in M9. We followed the worm mutagenesis protocol described by Brenner et al. (1974) with minor modifications. Every generation (~4 days), worms were washed from the plates with M9, collected in tubes, and resuspended in their respective mutagen solutions for 4 hours at 20 °C. After treatment, worms were transferred to fresh plates with NGM and *E. coli* OP50, ensuring maintenance of replicate identity and avoidance of mixing.

It is important to note that the population size of the worms was not altered at any point during the experiment. During both transfers—from the plates to the tubes and from the tubes back to the plates—all worms were transferred using 300 microliters of liquid. Specifically, M9 buffer was used for the transfer from plates to tubes, and the mutagen was used for the transfer from tubes back to plates. After the mutagen treatment, the liquid was carefully transferred onto the plate, causing parts of the *E. coli* colony to be dislodged. As a result, the worms would initially swim on the plate. We then left the plate open, allowing the liquid to be absorbed within 30-40 minutes, after which the worms resumed crawling on the plate.

525

526 *Male frequency assay*

527 To evaluate the impact of five generations of mutational exposure on male frequency, a male frequency
528 assay was conducted using both non-mutagenized (parent) worms and worms immediately after five
529 generations of mutagen treatment. Worms from each replicate population were bleached for age
530 synchronization. The lifetime progeny production of worms was monitored by placing a single L4 worm
531 on a 75mm x 13mm Petri plate, transferring it to a fresh plate every 24 hours. After 96 hours, the worm
532 was discarded, and the progeny produced over the four days were sexed and counted. Any worms that
533 died during the transfer process were excluded from the analysis**.** To ensure consistent sampling across
534 treatments, 30 replicate plates were prepared for each non-mutagenized parental group and 15 replicate
535 plates for each mutagenized population.

536

537 *Outcrossing frequency assay*

538 As with the male frequency assay, outcrossing frequency was assessed using the same set of replicate
539 plates — 30 for each non-mutagenized parental group and 15 for each mutagenized population. The
540 lifetime progeny production of each worm was monitored by placing a single L4 worm on a 75mm x 13mm
541 Petri plate, transferring both male and hermaphrodite to a fresh plate every 24 hours. After 96 hours,
542 both the male and hermaphrodite were discarded, and the progeny produced over the four days were
543 allowed to grow to adulthood.

544 Outcrossing in *C. elegans* is known to produce approximately 50% male progeny[49]. Based on male
545 frequency assays, we determined that spontaneous male production in the N2 and AB1 strains, both
546 before and after mutagenesis, was relatively low. Thus, plates for these strains were visually scanned for
547 males, and plates with male frequencies exceeding the expected spontaneous range (0.01–0.05) and
548 approaching 50% were categorized as outcrossing plates. For the CB4856 strain, which naturally exhibits
549 higher baseline male frequencies (15–17%), male frequency was calculated for each plate. A threshold of
550 25% males was set, and plates with male frequencies meeting or exceeding this value were marked as
551 outcrossing plates.

552 Outcrossing produces male and hermaphrodite offspring in a 1:1 ratio. Given the significantly low
553 male frequency in both the non-mutagenized and mutagenized N2 and AB1 strains, plates were scanned
554 for increased male frequency. If an unusually high number of males were observed, the mating pair was
555 considered to have outcrossed. In contrast, for the CB4856 strain, where the male frequency in non-

556 mutagenized parents was already around 20%, each progeny was sexed, and if the total male frequency
557 exceeded 40%, the mating pair was counted as an outcrossed pair.
558
559 *Relative fitness assay*
560 After five generations of mutagenesis, worms were frozen at -80°C and later recovered for relative fitness
561 assay at sixth generation- representing the fitness immediately after five generations of mutagenesis and
562 again after three generations of recovery from mutations using the standard *C. elegans* maintenance
563 protocol at 20 °C. During recovery, worms were maintained using the standard worm maintenance
564 protocol. For the relative fitness assay, one early-stage L1 worm from each focal strain (N2, CB4856, AB1)
565 was paired with one L1 worm from the non-mutagenized common competitor strain ST2 expressing green
566 fluorescent protein (GFP). Both strains were co-cultured on NGM plates with *E. coli* OP50 for eight days
567 (~2 generations) at 20 °C. Washed worms from replicate plates (11 per generation/mutant/strain
568 combination) were collected in M9 on a slide, imaged with and without a GFP filter on a fluorescent
569 microscope, and counted using a Python-based code specific for worm morphology (available on the
570 Github). Plates with zero worms of either type were excluded from analysis. Images captured without the
571 GFP filter counted all worms present on the plate- both the focal strain and the common competitor strain.
572 Conversely, images captured with the GFP filter only revealed the green, fluorescent competitor worms,
573 allowing for estimation of their abundance relative to the total worm population. With every relative
574 fitness assay of mutagenized and non-mutagenized worms an assay of relative fitness of non-mutagenized
575 control worm was also done.
576
577 *DNA extraction and Sequencing*
578 To assess the genomic impact of mutagenesis, high molecular weight DNA was extracted from worms
579 (after 3 generations of recovery) using the Cytiva kit (detailed procedure available in the Supplementary
580 Materials). Before starting out the DNA extraction we washed a minimum of ~10,0000 worms with M9
581 buffer from the plate into a tube and kept the worms rocking overnight to remove associated microbiome.
582 Pooled PacBio long-read sequencing was performed for parental worms of all three strains (N2, CB4856,
583 AB1) and one replicate per strain per mutagen. Additionally, Illumina short-read sequencing was
584 conducted for each replicate of each treatment (control, EMS, formaldehyde) for all strains. This strategy
585 enabled comprehensive analysis of genomic variants induced by mutagenesis and their phenotypic
586 consequences in subsequent generations.

587 For the comprehensive genomic analysis, we sequenced a total of 36 samples. Using PacBio long-
588 read sequencing, with an average depth of coverage of 94.52x, we sequenced nine samples: the three
589 parental strains (N2, CB4856, AB1) as controls, and six mutagenized replicates i.e., one replicate per strain
590 treated with each mutagen. Additionally, we performed Illumina short-read sequencing on 27 treatment
591 replicates, achieving an average depth of coverage of 72.11x. These 27 samples included the three
592 parental strains, 12 EMS-treated replicates, and 12 formaldehyde-treated replicates, with four replicates
593 per treatment for each strain.
594 The male frequency and outcrossing rate assays were conducted on worms derived from the same
595 batch that underwent five generations of mutagenesis. In contrast, the worms used for the relative fitness
596 assay and DNA extraction came from a different set, although they were mutagenized in the same manner
597 as the first batch.
598
599 **Data analysis**
600 Male frequency assay
601 Male frequency was calculated as the ratio of male progeny to the total progeny laid by each worm, and
602 this proportion was used as the unit of analysis. We assessed the normality of the data using the Shapiro-
603 Wilk test and by visualizing its distribution with a histogram. As expected, the distribution was right-
604 skewed, with a high proportion of zero values, indicating zero-inflation. Given the bounded nature (0 to
605 1) of the data and the excess zero values, we used a zero-inflated Beta regression for analysis. Since this
606 model requires proportion values strictly between 0 and 1, we applied an epsilon adjustment by adding a
607 small constant to all zero values to make the data more appropriate for the model. We performed the
608 analysis using the glmmTMB package in R (R version 4.1.0, released 2021-05-18)[50], implementing the
609 following model:
610 Male Frequency~Strain× Mutagen
611 In this model Strain (N2, AB1, CB4856) and Mutagen treatment (EMS, Formaldehyde and non-
612 mutagenized parents were included as fixed factors and an interaction term was added to check if the
613 effect of strain on male frequency varied by mutagen identity. This model accounts for both (1) the
614 continuous variation in male frequency and (2) the excess zero values separately, making it a better fit
615 than standard regression models. Further we did Tukey's HSD comparison [51] using emmeans package in R
616 [52].
617
618 *Relative fitness assay*

619 To quantify the relative fitness of the *C. elegans* strains N2, CB4856, and AB1, each plate was washed with
620 M9 buffer solution to prepare the worms for imaging. Two aliquots from the washed plate were then
621 placed onto a microscope slide and covered with a coverslip. We performed imaging to count the GFP-
622 expressing competitor strain and the non-GFP-expressing focal strain worms within these aliquots**.** The
623 allelic frequency of the non-fluorescent allele was calculated as the square root of proportion of non-GFP-
624 expressing worms to the total worm population. For statistical analysis, two-way ANOVAs were conducted
625 separately for each mutagen to assess the effects of strain and treatment duration on relative fitness.
626 Data normality was examined using the Shapiro-Wilk test and the residual distribution showed slight
627 variation from normality (Shapiro-Wilk test: $W = 0.9868$, $p < 0.01$), indicating variation from normality.
628 However, visual inspection of the residuals with histogram suggested an approximately normal
629 distribution. Since ANOVA is robust to minor deviations, we proceeded with ANOVA followed by pairwise
630 comparisons using Tukey's HSD test. The following linear model was applied:
631 anova(lm(RF ~ Strain*Generation, data = a)).
632 Following ANOVA, we performed pairwise comparisons using Tukey's Honest Significant
633 Difference (HSD) test to assess differences in relative fitness across strains and generations. We conducted
634 separate Tukey HSD tests within each strain to compare relative fitness before mutagenesis, after five
635 generations of mutagenesis, and after three generations of recovery.
636
637 *Bioinformatic analysis*
638 The bioinformatic analysis for identifying de novo mutations following PacBio sequencing involved a
639 comprehensive workflow to process and analyse long-read data, distinguishing mutations that arose in
640 the worms after mutagenesis from those existing in parental strains. Initially, raw PacBio reads were
641 aligned and indexed to reference genomes using minimap2 [53] and bcftools [54]; for the N2 and CB4856
642 strains, NCBI reference genomes were used, while the AB1 strain employed the recently annotated
643 genome of its descendant strain, SX3368, to account for genomic divergences within the lineage.
644 Subsequent variant calling was performed using Sniffles [55] for structural variants (SVs), such as insertions,
645 deletions, and inversions, capitalizing on the accuracy offered by long-read data.
646 For the specific identification of de novo mutations induced by mutagenesis, a comparative
647 analysis using bcftools contrasted the variant profiles of generation 8 worms (post three generations of
648 recovery) with their control parents, filtering out pre-existing genetic variations. Further quality control
649 and data filtering ensured only high-quality SNPs and SVs (quality score ≥ 20, depth ≥ 10) were retained.

650 We then used bedtools to provide insights into their potential impacts on exonic regions and regulatory
651 elements.
652 Additionally, similar bioinformatic steps were conducted with Illumina sequencing data.
653 Alignment and indexing were again performed on reference genomes, followed by SNP and SV calling
654 using Freebayes [56]and Manta[57], respectively. We initially called variants in pooled-population mode
655 (FreeBayes 'pooled') to account for minor alleles in the combined worm samples. However, when
656 intersecting these SNP calls with structural variants, we observed inflated counts due to multi-allelic sites
657 being represented multiple times. To avoid this over-counting artifact, we generated a separate VCF using
658 non-pooled calling parameters and used these calls for all subsequent intersection and annotation
659 analyses. Enhanced SV detection was achieved with VGtoolkit [58] through graph-based variant calling,
660 which was then integrated with SVs identified by Sniffles 2.0 using Survivor 1.0.3[59], creating a
661 comprehensive view of the genomic alterations induced by the mutagens.
662 Long-read sequencing was performed on the three non-mutagenized parent strains and one
663 replicate per strain for each mutagen treatment. Short-read sequencing covered all replicates across all
664 strains. For structural variation analysis, we merged variants identified by both long-read and short-read
665 callers using Survivor 1.0.3. For SNP analysis, we relied solely on variants identified by short-read callers.
666
667 Annotation of Transposable Elements and Intersection with Structural Variants
668 Transposon annotation is facilitated in complete and contiguous sequences, thus annotation of TEs was
669 done in the N2, AB, and CB reference genomes rather than directly from the raw sequence data. We used
670 the TransposonUltimate reasonaTE[60] pipeline for TE annotation in each of the reference genomes. Briefly,
671 TEUltimate reasonaTE wraps 13 previously published TE annotation tools, which include: RepeatMasker
672 v4.1.1 **,** RepeatModeler v2.0.1, SINE-Scan , SINE-Finder, LTRPred , LTRHarvest, TIRvish , MUSTv2 , MITE-
673 Tracker, MITEFinderII**,** HelitronScanner, TransposonPSI for transposon discovery. While some functions
674 are optional, all subprograms were run with default parameters in this study. The resulting annotations
675 are filtered, merged, and clustered using CD-HIT v4.8.1 [61] and BLASTN v2.10.1 [62]. The final annotations
676 from FinalAnnotations_Transposons.gff3 in the final Results folder were used for downstream analysis for
677 each of the three focal strains.
678 Evidence of transposon movement was assumed from structural variants which overlapped with
679 transposon annotations. The TEUltimatedeTEct pipeline [60] was used for intersection of transposon and
680 structural variant annotations. De novo VCF structural variant files, as generated above, were intersected
681 with transposon annotations and reported as an event if overlapping by at least 10% and lengths were

682 similar by at least 50%. Finally, any annotation smaller than 50 base pairs, or larger than 1% of the genome
683 was discarded. Plots were generated from the transpositionEvents.gff3 file in RStudio with R v4.3.2 (R
684 Core Team, 2023).
685
686 Literature Cited


687 1. Lynch, M. *et al*. *The Mutational Meltdown in Asexual Populations*. *We Are Very*
688 *Grateful To*. *Journal of Heredity* vol. 84 (1993).

689 2. Lynch, M. *et al*. Perspective: Spontaneous deleterious mutation. *Evolution (N Y)* **53**,
690 645–663 (1999).

691 3. Robert, F. & Pelletier, J. Exploring the Impact of Single-Nucleotide Polymorphisms on
692 Translation. *Frontiers in Genetics* vol. 9 Preprint at
693 https://doi.org/10.3389/fgene.2018.00507 (2018).

694 4. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome.
695 *Nature Reviews Genetics* vol. 7 85–97 Preprint at https://doi.org/10.1038/nrg1767
696 (2006).

697 5. Collins, R. L. *et al.* A structural variation reference for medical and population
698 genetics. *Nature* **581**, 444–451 (2020).

699 6. Mahmoud, M. *et al*. Structural variant calling: The long and the short of it. *Genome*
700 *Biology* vol. 20 Preprint at https://doi.org/10.1186/s13059-019-1828-7 (2019).

701 7. Ho, S. S., Urban, A. E. & Mills, R. E. Structural variation in the sequencing era. *Nature*
702 *Reviews Genetics* vol. 21 171–189 Preprint at https://doi.org/10.1038/s41576-019-
703 0180-9 (2020).

704 8. Glémin, S. How are deleterious mutations purged? Drift versus nonrandom mating.
705 *Evolution (N Y)* **57**, 2678–2687 (2003).

706 9. Whitlock, M. C. & Agrawal, A. F. Purging the genome with sexual selection: Reducing
707 mutation load through selection on males. *Evolution* vol. 63 569–582 Preprint at
708 https://doi.org/10.1111/j.1558-5646.2008.00558.x (2009).

709 10. García-Dorado, A. Understanding and predicting the fitness decline of shrunk
710 populations: Inbreeding, purging, mutation, and standard selection. *Genetics* **190**,
711 1461–1476 (2012).

11. Byers, D. L. & Waller, D. M. *Do Plant Populations Purge Their Genetic Load? Effects of Population Size and Mating History on Inbreeding Depression*. *Source: Annual Review of Ecology and Systematics* vol. 30 https://about.jstor.org/terms (1999).

12. Szövényi, P. *et al*. Efficient purging of deleteriousmutations in plants with haploid selfing. *Genome Biol Evol* **6**, 1238–1252 (2014).

13. MacPherson, B., Scott, R. & Gras, R. Sex and recombination purge the genome of deleterious alleles: An Individual Based Modeling Approach. *Ecological Complexity* **45**, (2021).

14. Sianta, S. A., Peischl, S., Moeller, D. A. & Brandvain, Y. The efficacy of selection may increase or decrease with selfing depending upon the recombination environment. *Evolution (N Y)* **77**, 394–408 (2023).

15. Hartfield, M. Evolutionary genetic consequences of facultative sex and outcrossing. *J Evol Biol* **29**, 5–22 (2016).

16. Hoffmann, A. A. & Rieseberg, L. H. Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics* vol. 39 21–42 Preprint at https://doi.org/10.1146/annurev.ecolsys.39.110707.173532 (2008).

17. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).

18. Gaertner, B. E. & Phillips, P. C. Caenorhabditis elegans as a platform for molecular quantitative genetics and the systems biology of natural variation. *Genet Res (Camb)* **92**, 331–348 (2010).

19. Kutscher, L. M. & Shaham, S. Forward and reverse mutagenesis in C. elegans. *WormBook : the online review of C. elegans biology* 1–26 Preprint at https://doi.org/10.1895/wormbook.1.167.1 (2014).

20. Kondrashov, A. S. Selection against harmful mutations in large sexual and asexual populations. *Genet Res* **40**, 325–332 (1982).

21. Charlesworth, B. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet Res* **55**, 199–221 (1990).

22. Katju, V., LaBeau, E. M., Lipinski, K. J. & Bergthorsson, U. Sex change by gene conversion in a Caenorhabditis elegans fog-2 mutant. *Genetics* **180**, 669–672 (2008).

743 23. Racimo, F. & Schraiber, J. G. Approximation to the Distribution of Fitness Effects
744 across Functional Categories in Human Segregating Polymorphisms. *PLoS Genet* **10**,
745 (2014).

746 24. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new
747 mutations. *Nature Reviews Genetics* vol. 8 610–618 Preprint at
748 https://doi.org/10.1038/nrg2146 (2007).

749 25. Keightley, P. D. & Eyre-Walker, A. Joint inference of the distribution of fitness effects
750 of deleterious mutations and population demography based on nucleotide
751 polymorphism frequencies. *Genetics* **177**, 2251–2261 (2007).

752 26. Haller, B. C. & Messer, P. W. SLiM 4: Multispecies Eco-Evolutionary Modeling.
753 *American Naturalist* **201**, E127–E139 (2023).

754 27. Ospina, R. & Ferrari, S. L. P. On Bias Correction in a Class of Inflated Beta Regression
755 Models. *Int J Stat Probab* **1**, (2012).

756 28. Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdisciplinary*
757 *Reviews: Computational Statistics* vol. 2 433–459 Preprint at
758 https://doi.org/10.1002/wics.101 (2010).

759 29. Riehl, K., Riccio, C., Miska, E. A. & Hemberg, M. TransposonUltimate: software for
760 transposon classification, annotation and detection. *Nucleic Acids Res* **50**, E64
761 (2022).

762 30. Estes, S., Phillips, P. C., Denver, D. R., Thomas, W. K. & Lynch, M. *Mutation*
763 *Accumulation in Populations of Varying Size: The Distribution of Mutational Effects*
764 *for Fitness Correlates in Caenorhabditis Elegans*.

765 31. Konrad, A. *et al.* Mitochondrial mutation rate, spectrum and heteroplasmy in
766 Caenorhabditis elegans spontaneous mutation accumulation lines of differing
767 population size. *Mol Biol Evol* **34**, 1319–1334 (2017).

768 32. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with
769 Knives. *Molecular Cell* vol. 40 179–204 Preprint at
770 https://doi.org/10.1016/j.molcel.2010.09.019 (2010).

771 33. Limpose, K. L., Corbett, A. H. & Doetsch, P. W. BERing the burden of damage:
772 Pathway crosstalk and posttranslational modification of base excision repair
773 proteins regulate DNA damage management. *DNA Repair* vol. 56 51–64 Preprint at
774 https://doi.org/10.1016/j.dnarep.2017.06.007 (2017).

775 34. Whitlock, M. C. & Agrawal, A. F. Purging the genome with sexual selection: Reducing
776 mutation load through selection on males. *Evolution* vol. 63 569–582 Preprint at
777 https://doi.org/10.1111/j.1558-5646.2008.00558.x (2009).

778 35. Artieri, C. G., Haerty, W., Gupta, B. P. & Singh, R. S. Sexual selection and
779 maintenance of sex: Evidence from comparisons of rates of genomic accumulation
780 of mutations and divergence of sex-related genes in sexual and hermaphroditic
781 species of Caenorhabditis. *Mol Biol Evol* **25**, 972–979 (2008).

782 36. Negi, P., Rai, A. N. & Suprasanna, P. Moving through the stressed genome: Emerging
783 regulatory roles for transposons in plant stress response. *Frontiers in Plant Science*
784 vol. 7 Preprint at https://doi.org/10.3389/fpls.2016.01448 (2016).

785 37. Wright, S. I. & Schoen, D. J. *Transposon Dynamics and the Breeding System*.
786 *Genetica* vol. 107 (1999).

787 38. Dolgin, E. S. & Charlesworth, B. The fate of transposable elements in asexual
788 populations. *Genetics* **174**, 817–827 (2006).

789 39. Arkhipova, I. & Meselson, M. *Transposable Elements in Sexual and Ancient Asexual*
790 *Taxa*. www.pnas.org.

791 40. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: A trade-
792 off between reduced transposition and deleterious effects on neighboring gene
793 expression. *Genome Res* **19**, 1419–1428 (2009).

794 41. Sterken, M. G., Snoek, L. B., Kammenga, J. E. & Andersen, E. C. The laboratory
795 domestication of Caenorhabditis elegans. *Trends in Genetics* vol. 31 224–231
796 Preprint at https://doi.org/10.1016/j.tig.2015.02.009 (2015).

797 42. Thompson, O. A. *et al.* Remarkably divergent regions punctuate the genome
798 assembly of the Caenorhabditis elegans hawaiian strain CB4856. *Genetics* **200**,
799 975–989 (2015).

800 43. Anderson, J. L., Morran, L. T. & Phillips, P. C. Outcrossing and the maintenance of
801 males within C. elegans populations. in *Journal of Heredity* vol. 101 (2010).

802 44. Teotonio, H., Carvalho, S., Manoel, D., Roque, M. & Chelo, I. M. Evolution of
803 outcrossing in experimental populations of caenorhabditis elegans. *PLoS One* **7**,
804 (2012).

805 45. Charlesworth, B. & Charlesworth, D. The population dynamics of transposable
806 elements. *Genet Res* **42**, 1–27 (1983).

807 46. Charlesworth, B. OPTIMIZATION MODELS, QUANTITATIVE GENETICS, AND
808 MUTATION. *Evolution (N Y)* **44**, 520–538 (1990).

809 47. Otto, S. P. Unravelling the evolutionary advantage of sex: A commentary on
810 'Mutation-selection balance and the evolutionary advantage of sex and
811 recombination' by Brian Charlesworth. *Genetics Research* vol. 89 447–449 Preprint
812 at https://doi.org/10.1017/S001667230800966X (2008).

813 48. Haller, B. C. & Messer, P. W. SLiM 4: Multispecies Eco-Evolutionary Modeling.
814 *American Naturalist* **201**, E127–E139 (2023).

815 49. Yin, D. & Haag, E. S. Evolution of sex ratio through gene loss. *Proc Natl Acad Sci U S*
816 *A* **116**, 12919–12924 (2019).

817 50. Bolker, B. *Getting Started with the GlmmTMB Package*. (2022).

818 51. Abdi, H. & Williams, L. J. *Encyclopedia of Research Design*.
819 http://www.utd.edu/~herve.

820 52. Searle, S. R., Speed, F. M. & Milliken, G. A. Population marginal means in the linear
821 model: An alternative to least squares means. *American Statistician* **34**, 216–221
822 (1980).

823 53. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**,
824 3094–3100 (2018).

825 54. Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

826 55. Smolka, M. *et al*. Detection of mosaic and population-level structural variants with
827 Sniffles2. *Nat Biotechnol* (2024) doi:10.1038/s41587-023-02024-y.

828 56. ena Ypsilanti, A. R. *et al*. Transcriptional network orchestrating regional patterning of
829 cortical progenitors. doi:10.1073/pnas.2024795118/-/DCSupplemental.

830 57. Chen, X. *et al*. Manta: Rapid detection of structural variants and indels for germline
831 and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

832 58. Hickey, G. *et al*. Genotyping structural variants in pangenome graphs using the vg
833 toolkit. *Genome Biol* **21**, (2020).

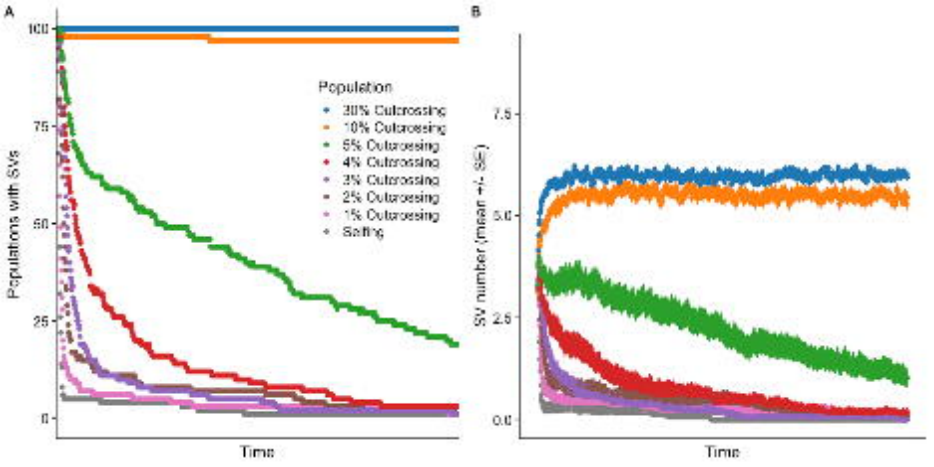
834 59. Jeffares, D. C. *et al*. Transient structural variations have strong effects on
835 quantitative traits and reproductive isolation in fission yeast. *Nat Commun* **8**, (2017).

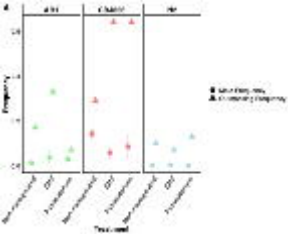
836 60. Riehl, K., Riccio, C., Miska, E. A. & Hemberg, M. TransposonUltimate: software for
837 transposon classification, annotation and detection. *Nucleic Acids Res* **50**, E64
838 (2022).

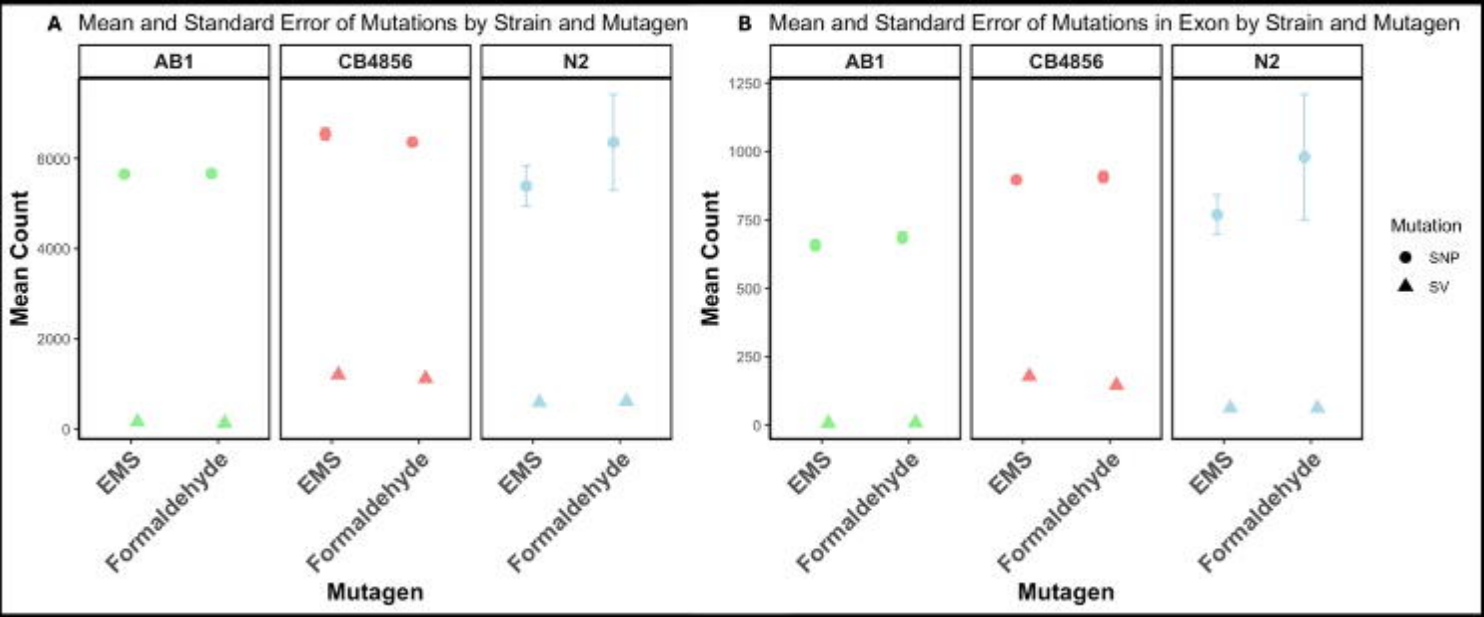
839 61. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-
840 generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

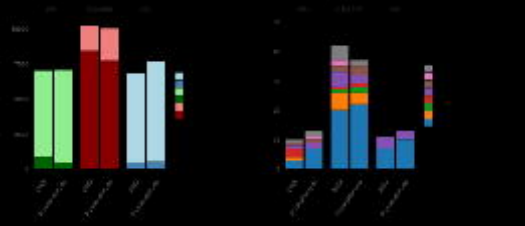
841 62. Camacho, C. *et al*. BLAST+: Architecture and applications. *BMC Bioinformatics* **10**,
842 (2009).

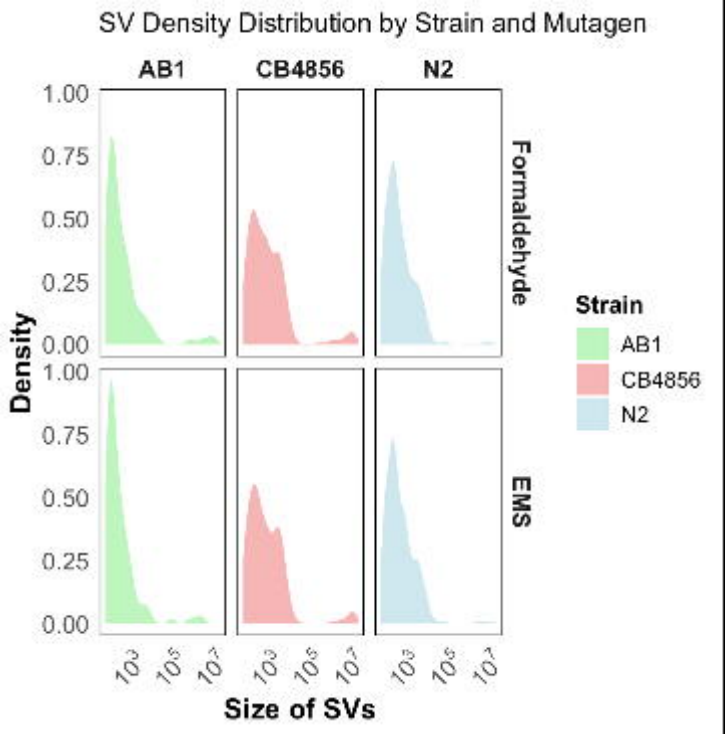
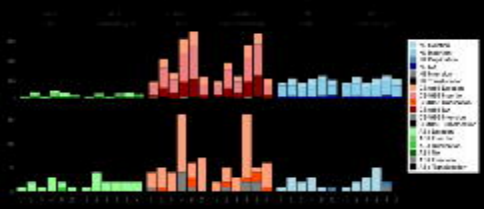
843


A
B
Population
30% Outcrossing
10% Outcrossing
5% Outcrossing
4% Outcrossing
3% Outcrossing
2% Outcrossing
1% Outcrossing
Selfing
Populations with SVs
SV number (mean +/- SE)
Time
Time
100
75
50
25
0
7.5
5.0
2.5
0.0




A
Mean and Standard Error of Mutations by Strain and Mutagen
AB1
CB4856
N2
Mean Count
8000
4000
2000
0
EMS
Formaldehyde
Mutagen
B
Mean and Standard Error of Mutations in Exon by Strain and Mutagen
AB1
CB4856
N2
Mean Count
1250
1000
750
500
250
0
EMS
Formaldehyde
Mutagen
Mutation
SNP
SV




SV Density Distribution by Strain and Mutagen
AB1
CB4856
N2
Formaldehyde
EMS
Density
1.00
0.75
0.50
0.25
0.00
Size of SVs
Strain
AB1
CB4856
N2