Evolution of the rate, spectrum, and fitness effects of mutation under minimal selection in *Caenorhabditis elegans*

Sayran Saber[1,*], Md. Monjurul Islam Rifat[1,*], Fahimeh Rahimi[1], Michael Snyder[1], Amber Singh[1], Benjamin Eickwort[1], Yell Newhall[1], Moein Rajaei[1,4], Ayush Shekhar Saxena[1,5], Robyn E. Tanny[2], Vaishali Katju[3], Erik C. Andersen[2], Juannan Zhou[1], and Charles F. Baer[1,6,†]

1- Department of Biology, University of Florida, Gainesville, FL USA

2- Department of Biology, Johns Hopkins University, Baltimore, MD USA

3- Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

4- Current Address: School of Public Health, Yale University, New Haven, CT USA

5- Current Address: Regeneron, Inc., White Plains, NY USA

6- University of Florida Genetics Institute, Gainesville, FL USA

* These authors contributed equally

[†] Correspondence to: cbaer@ufl.edu

Keywords: Cost of fidelity, Distribution of fitness effects, Drift barrier, Epistasis, Mutation accumulation, Mutation rate, Mutation spectrum, Mutational bias

Abstract

Evolution of mutation rate is of longstanding interest in evolutionary biology. The classical explanation is a “cost of fidelity” (CoF), which posits that the mutation rate is under stabilizing selection resulting from a tradeoff between selection to reduce deleterious mutations and selection to minimize resources devoted to genome surveillance. Alternatively, the “drift barrier” (DB) hypothesis posits that mutation rate is always under directional selection downward; at equilibrium, the increase in fitness from a further increase in replication fidelity is too small to overcome genetic drift and an (assumed) mutational bias to increase mutation rate.

To test these alternative hypotheses, we propagated *C. elegans* “mutation accumulation” (MA) lines under minimal selection for ~150 generations, divided each MA line into two sublines and propagated them for another ~150 generations. The base-substitution mutation rate did not change, as predicted by the CoF. Conversely, the insertion/deletion rate increased, as predicted by the DB. These findings add to the evidence that the indel mutation rate is more sensitive to genomic stress than is the nucleotide substitution rate.

A related issue concerns the nature of selection on deleterious mutations. If deleterious mutations interact synergistically, several evolutionary quandaries could be resolved. To quantify mutational effects on fitness we assayed competitive fitness of the MA lines at 150 and 300 generations. Mutations had a mean effect of -0.1%, and a model of constant mutational effects (no epistasis) satisfactorily explains the results. There was a weak negative association between mutation rate and fitness, as predicted by the DB hypothesis.

The evolution of the mutation rate is of longstanding interest in evolutionary biology. The central conundrum is: since most mutations with discernible effects on fitness are deleterious, why does the mutation rate not evolve to zero (Fisher 1930; Sturtevant 1937)? The two most compelling hypotheses are, in historical chronological order, (1) the Cost of Fidelity (CoF; Kimura 1967) and (2) the Drift Barrier (DB; Lynch 2008). The CoF posits that the optimum mutation rate represents a tradeoff between direct selection to minimize the input of deleterious mutations and indirect selection to minimize the resources (time, energy, molecules) devoted to genome surveillance, resulting in stabilizing selection on mutation rate (Figure 1A). The DB posits that directional selection acts to reduce the mutation rate until the fitness increase resulting from a further reduction of the mutation rate is less than the reciprocal of the effective population size (N_e), at which point selection is too weak to counteract genetic drift and the (assumed) mutational bias toward increased mutation rate (Figure 1B).

The DB hypothesis predicts that a decrease(increase) in N_e should result in an increase(decrease) in mutation rate. The DB does not preclude the occasional transient decrease in mutation rate due to the occurrence of antimutator alleles, just that selection is directional over the long term and the equilibrium is established by the drift barrier. The DB hypothesis also assumes, at least implicitly, that the fitness function is concave downward in the vicinity of the drift barrier (Figure 1B), i.e. "*…a point must eventually be reached where the advantage of any further increase in replication fidelity is smaller than the power of random genetic drift.*" (Lynch 2011).

The CoF makes an analogous prediction: if the mutation rate is at the optimum, a change in N_e, should not result in a directional change in the mutation rate. However, populations at the same optimum should experience a predictable change in the variance whereby a decrease(increase) in N_e should result in increased(decreased) variance around the optimum. Stabilizing selection comes with its own implicit drift barrier because the fitness function is concave downward; mutations with fitness effects $s<1/N_e$ will accumulate in the

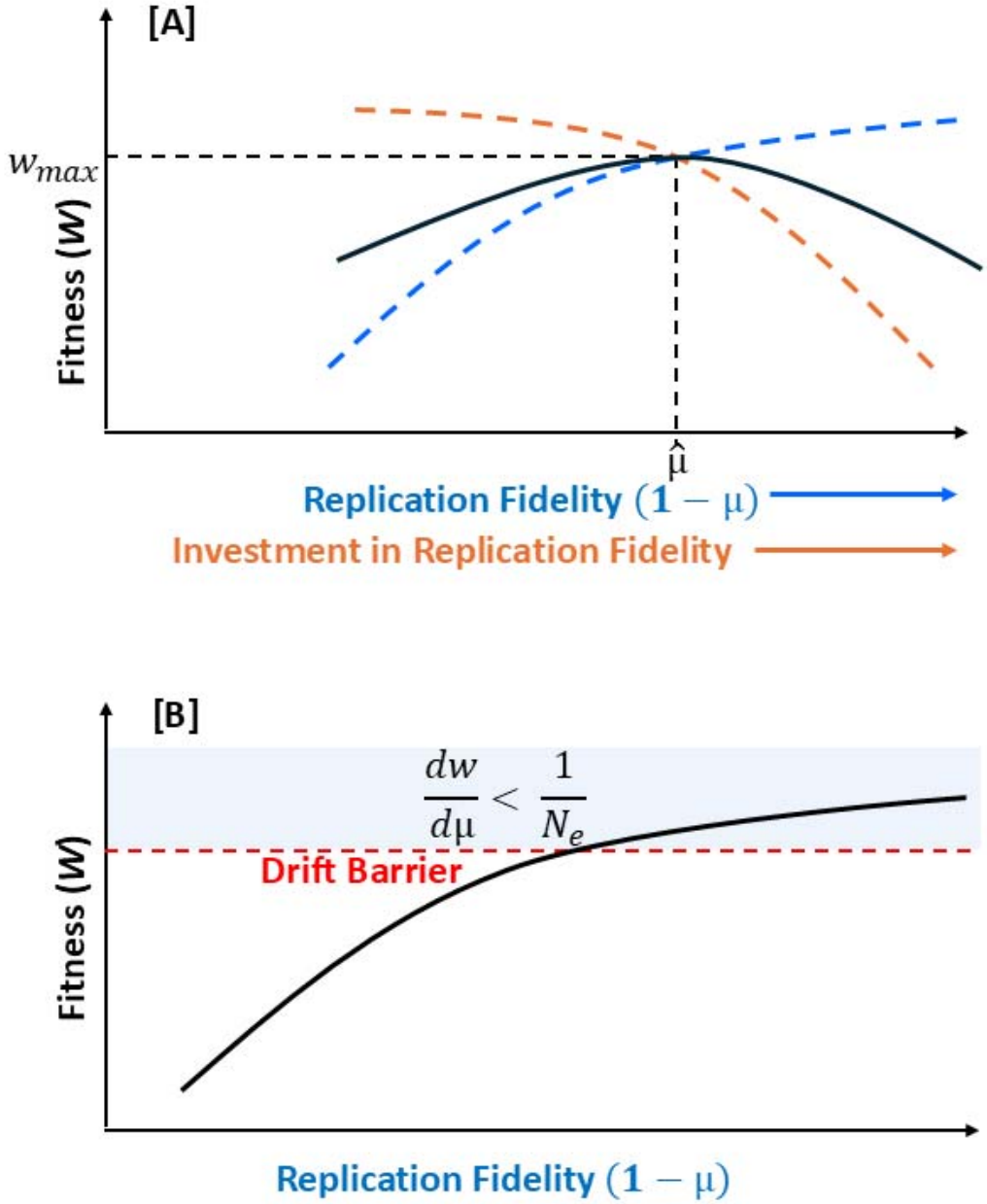


Figure 1. Theoretical models of mutation rate evolution. (**A**) Cost of Fidelity. (**B**) Drift Barrier.

population, irrespective of whether they increase or decrease trait value. However, the prediction of no directional change under relaxed selection is predicated on symmetry of the fitness function around the optimum (Figure 1A). If the fitness function is asymmetric, a population at mutation-selection-drift equilibrium will exhibit a mutational bias, such that more mutations cause the mutation rate to change in the direction of stronger selection than in the direction of weaker selection (Supplementary Figure 1). Without a mutational bias, the population would not be at equilibrium and would be under directional selection.

A second feature of the mutational process of longstanding interest is the distribution of fitness effects (DFE) of new mutations (Fisher 1930; Charlesworth 1996; Eyre-Walker and Keightley 2007). Of particular interest is how deleterious mutations interact. Whether the interaction between deleterious mutations is multiplicative (no epistasis), synergistic (negative epistasis) or diminishing-returns (positive epistasis) has important implications for evolution, prominently including the evolution of sex and recombination (Barton 1995; Otto 2009).

To investigate these questions in a tractable experimental system, we employed a "second order mutation accumulation" (O2MA) experiment (Ávila, et al. 2006; Saxena, et al. 2019; Liu and Zhang 2021) using the nematode *C. elegans* (Figure 2A). We initially constructed a set of "first order MA" (O1MA) lines from a nearly isogenic progenitor, using the standard *C. elegans* MA protocol (Keightley and Caballero 1997; Vassilieva and Lynch 1999; Baer, et al. 2005). After ~150 generations, each O1MA line was subdivided into two replicate O2MA lines and propagated for an additional ~150 generations, at which time the genomes of the O2MA lines were sequenced, and competitive fitness of the O1MA and O2MA lines was assayed (Figure 2B). The sequence data allow us to quantify the evolution of the rate and spectrum of mutation over ~300 generations of minimal selection. The fitness data allow us to quantify the evolution of the mean and variance of fitness effects of the accumulated mutations.

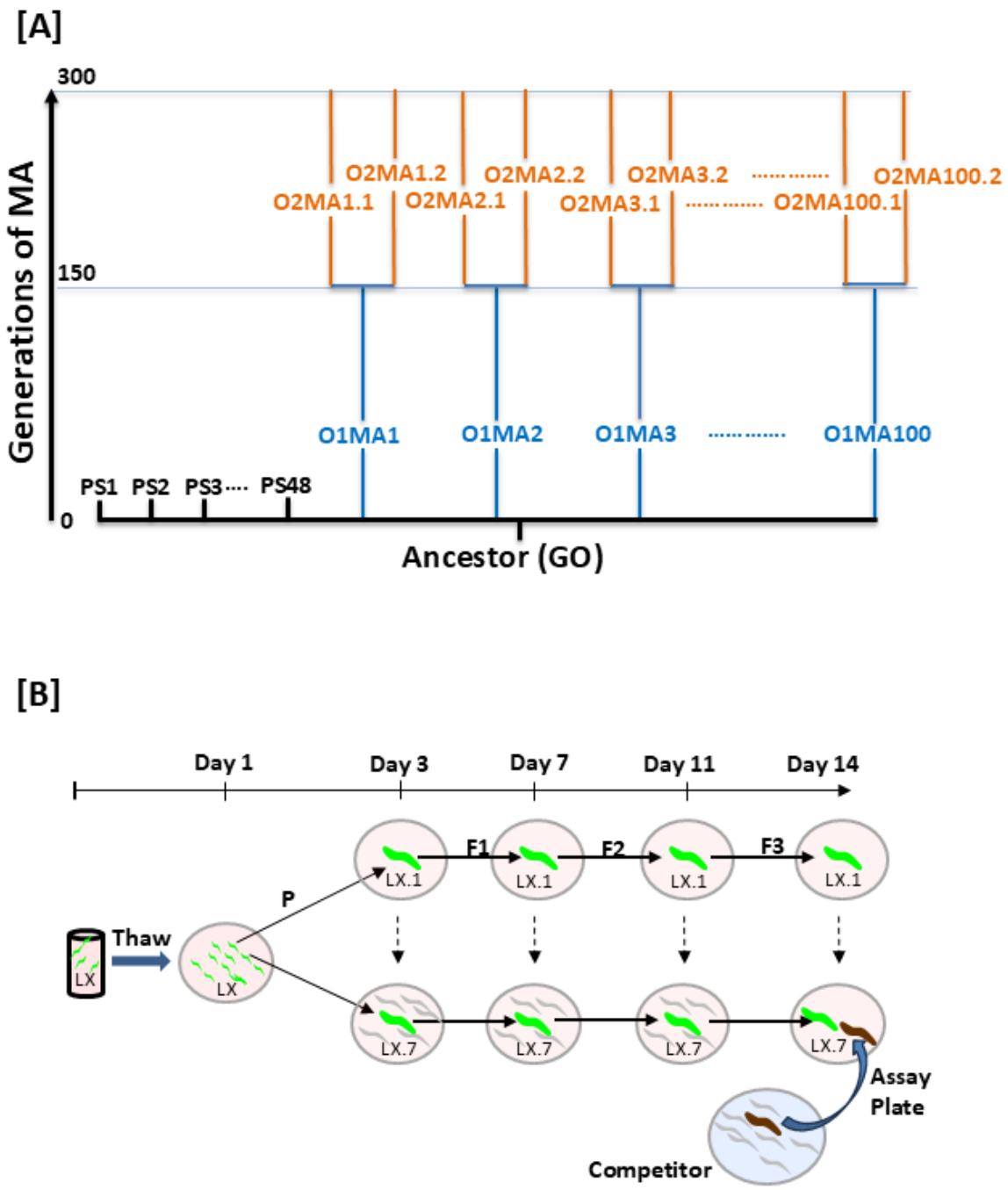


Figure 2. (**A**) Schematic depiction of the MA experiment. Short black lines depict ancestral G0 "pseudolines" (PS), blue lines depict O1MA lines, orange lines depict O2MA lines. (**B**) Schematic depiction of the competitive fitness assay. Label "LX" refers to "Line X" (MA or G0), label LX.1,…LX.7 refers to replicates 1-7 of a line in the assay. See Methods for details.

Results

Mutation rate. Over the first ~150 generations we observed 2938 putative mutations (2430 SNVs, 348 deletions, and 160 insertions (Supplementary Tables 1 and 2; Supplementary Figure 2). Ignoring false inferences for the moment, the first-order per-nucleotide SNV rate $\mu_{O1,SNV}$=1.77±0.038×10^{-9}/generation; the indel rate $\mu_{O1,Indel}$=0.36±0.019×10^{-9}/generation (Table 1). The total first-order mutation rate (μ_{O1}) is 2.14±0.042×10^{-9}/generation. In the second ~150 generations of the experiment we observed 6631 total variants (5084 SNVs, 1016 deletions, 763 insertions). Still ignoring false inferences, the second-order SNV rate $\mu_{O2,SNV}$=1.90±0.034×10^{-9}/generation. The second-order indel rate $\mu_{O2,Indel}$=0.60±0.024×10^{-9}/generation. The total second-order mutation rate μ_{O2}= 2.48±0.046×10^{-9}/generation.

False positives (FP) and false negatives (FN) affect the estimated mutation rates in predictable ways (Supplementary Figure 3). FPs artificially inflate the inferred μ_{O2}, and false negatives (FN) artificially deflate μ_{O2} on average. Ignoring FNs results in a conservative test of the hypothesis that $\mu_{O2} > \mu_{O1}$. Because we have only a rough estimate of the number of FPs (≈ 2 SNVs and 0.33 indels per genome; Supplementary Table 3), we tested the null hypothesis that $\mu_{O2} = \mu_{O1}$ by "titrating" the number of FPs by simulation and asking the question: what is the fewest FPs per O2MA line that would lead us to falsely reject the null hypothesis $\mu_{O2} = \mu_{O1}$ (i.e., Type I error)? For SNVs, there must be fewer than 0.6 FPs for the observed difference $\Delta\mu_{SNV}$ = 0.23×10^{-9}/generation to not be sufficiently explained by FPs (Figure 3A), so we conclude that there is no evidence that $\mu_{O2,SNV} > \mu_{O1,SNV}$. For indels, the same test shows the critical number of FPs per line is 2.7 (Figure 3B). The probability of observing no more than one FPs in three replicates (see Methods and Supplementary Table 3) if the true Poisson mean is 2.7 is approximately 0.004, so we conclude that the evidence suggests that $\mu_{O2,Indel} > \mu_{O1,Indel}$.

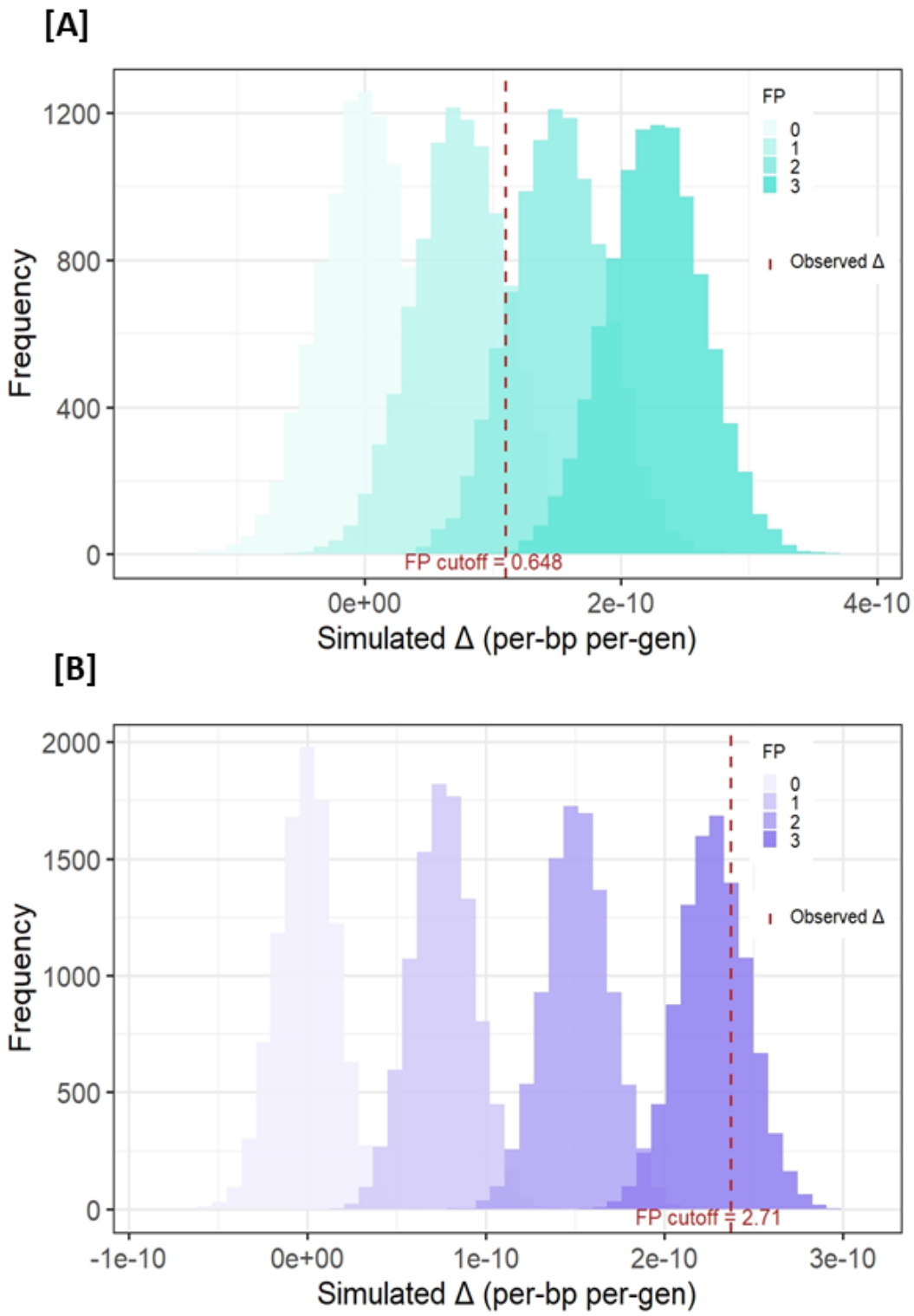


Figure 3. Distribution of simulated values of $\Delta\mu=\mu_{O2}-\mu_{O1}$ with increasing number of false positives. Red dashed line is the observed value. "FP cutoff" is the minimum number of false positives necessary for the data to be consistent with the hypothesis $\Delta\mu=0$. (**A**) SNVs (**B**) Indels.

Some types of mutations occur independently of genome replication and accrue at an approximately constant rate in chronological time. If the chronological mutation rate remains constant but the generation time increases, the result will be an increase in the per-generation mutation rate. There was little difference in the average number of generations between the O1 and O2 lines over the course of the experiment (O1 mean = 145.6 generations; O2 mean = 144.2 generations), and the conclusions drawn from mutation rates scaled per-generation apply similarly to mutation rates scaled per-day (Table 1). The number of indels was positively correlated with generations of MA in both O1 ($r_{O1,indel}$=0.29, p<0.004) and O2 lines ($r_{O2,indel}$=0.25, p<0.001). In the O1 lines, the number of SNVs was uncorrelated with the number of generations ($r_{O1,SNV}$=0.10, p>0.32), whereas in the O2 lines the number of SNVs was positively correlated with the number of generations ($r_{O2,SNV}$=0.46, p<<0.0001). The lack of correlation of number of SNVs in the O1MA lines is likely due to lack of power resulting from the smaller sample size and the lack of variation in generations of MA. Taken together, we reject the hypothesis that mutation rate is independent of generation time, but we cannot rule out a contribution of replication-independent mutation.

Mutation spectrum. Consistent with the observed lack of evolution of μ_{SNV}, the SNV mutation spectra did not differ between O1MA and O2MA lines (Monte Carlo Fisher's Exact Test, p > 0.12) (Figure 4A,B). G:C→A:T transitions (28.6%) and A:T→T:A transversions (28.1%) were the most common classes of base substitutions, as is typical in *C. elegans* MA experiments(Konrad, et al. 2019; Rajaei, et al. 2021).

The indel spectrum, in contrast, was significantly different between O1MA and O2MA lines (Figure 4C,D). In the O1MA lines, deletions outnumbered insertions by nearly 2.2:1, whereas the deletion bias in the O2MA lines was significantly reduced to ~1.6:1 (Fisher's Exact Test (FET), p < 0.0001). Considered separately, the deletion spectrum differed significantly between O1 and O2 lines (FET, p < 0.003), with an increase in the shortest (1 bp) and longest

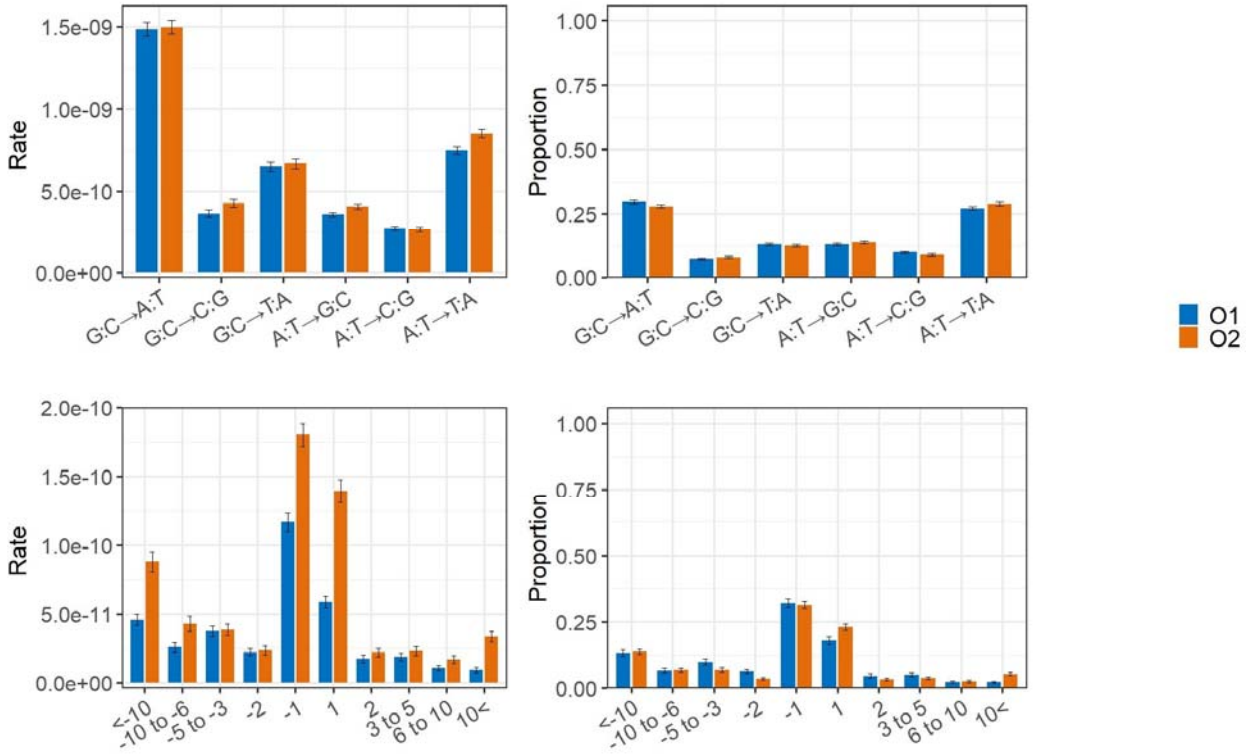


Figure 4. Type-specific mutation rates and spectra. (**A**) SNV mutation rates (**B**) SNV frequency spectrum (**C**) Indel mutation rates (**D**) Indel frequency spectrum.

(>10 bp) categories and concomitant decreases in the intermediate categories. The insertion spectrum also differed between O1 and O2 (FET, $p < 0.0001$), also with increases in the 1 bp and >10 bp categories and decreases in the intermediate categories.

Variance in mutation rate. To begin, we ask: can the observed variance in mutation rate among O1MA lines ($V_{\mu O1}$) be sufficiently explained by sampling variance of a uniform mutational process? For SNVs the answer is yes; $V_{\mu O1,SNV}$ is sufficiently explained by sampling if mutation conforms to a uniform Poisson process ($p>0.09$, Supplementary Figure 4A). For indels, however, $V_{\mu O1,Indel}$ is too great to be explained by a uniform Poisson process ($p<0.02$, Supplementary Figure 4B).

To extend the analysis to the O2MA lines, FPs must be taken into account. We assume a liberal average of 3 SNV and 1 indel FPs per sample, leading to a conservative test of the hypothesis of a uniform mutational process. For both SNVs and indels there is more variance among O2MA lines than is consistent with a uniform mutational process (Supplementary Figure 5).

A different approach is to ask: how much heritable variation in μ_{O2} exists among O1MA families? Over a given time period, a trait may not be heritable either because it evolves too slowly for significant variation to accumulate, or because it evolves so rapidly that the signal of ancestor-descendant relationship is erased.

Neither $\mu_{O2,SNV}$ nor $\mu_{O2,Indel}$ is significantly heritable ($h^2_{\mu_{O2,SNV}} = 0.037$, p>0.71; $h^2_{\mu_{O2,Indel}} = 0.089$, p>0.38), nor is total mutation rate ($h^2_{\mu_{O2}} = 0.046$, p>0.65). For SNVs, the lack of heritability is consistent with inferred uniformity of the O1MA process. For indels, the simplest explanation is that (*i*) heritability is inherently low, and (*ii*) FPs add additional noise, so 96 O1 families provide insufficient power to detect heritability on the order of 10% over 150 generations. A more speculative possibility is that in the absence of purifying selection, indel mutation rate is an accelerating function of mutation load (Agrawal 2002; Shaw and Baer 2011).

Mutational effects on fitness. Because all mutations in an MA line are in complete linkage disequilibrium, the effects of individual mutations are unidentifiable. However, knowing the number of mutations in each MA line allows a straightforward estimate of the first two moments of the DFE, because the sum of the effects of a large number of individual mutations within a line can be well-approximated by a normal distribution regardless of the underlying DFE, according to the Central Limit Theorem.

Competitive fitness (*W*) decreased by about 0.2% per generation over the first 150 generations of MA and by about half that over the next 150 generations of MA (Supplementary

Figure 5, Supplementary Tables 4 and 5). There is a weak negative association between mutation rate and fitness (Supplementary Figure 6).

Given that the number of mutations is similar between O1MA and O2MA lines, the implication is that the mean effect on competitive fitness of O1 mutations (u_{O1}) is greater than that of O2 mutations (u_{O2}). To test that hypothesis, we compared a model in which the mean mutational effect u is constrained to remain constant over the entire 300 generations ("uniform effects") to a "two-stage" model in which u_{O1} and u_{O2} are allowed to differ. Models were compared using the Widely Applicable Information Criterion (WAIC) and leave-one-out cross-validation (LOO) to assess model fit and predictive performance (Vehtari, et al. 2017). The differences between models are substantial (Δ_{WAIC} = 13.52; Δ_{LOO} = 12.69) and favor the uniform effects model. The point estimate of the mean mutational effect u = -0.0139, with variance $\sigma_g^2 = 1.22\times10^{-4}$ (Figure 5).

Discussion

Our primary goal was to test empirically the predictions of the two prominent alternative theoretical models of the evolution of mutation rate. The Drift Barrier (DB) model predicts that the mutation rate should increase when selection is relaxed. Explicit in the DB model is the existence of a mutational bias in the direction of increased mutation rate. The Cost of Fidelity (CoF) model posits that mutation rate is under stabilizing selection, whereby direct selection to reduce the input of deleterious mutations is counterbalanced by indirect selection to minimize the resources devoted to genome surveillance. The "strong" form of the CoF is predicated on symmetry of the fitness function around the optimum and predicts that relaxation of selection will not lead to a change in the mean mutation rate, but that the variance in mutation rate will increase. The "weak" form of the CoF, which relaxes the assumption of symmetry of the fitness function, also predicts an increase in the variance in mutation rate upon relaxation of selection.

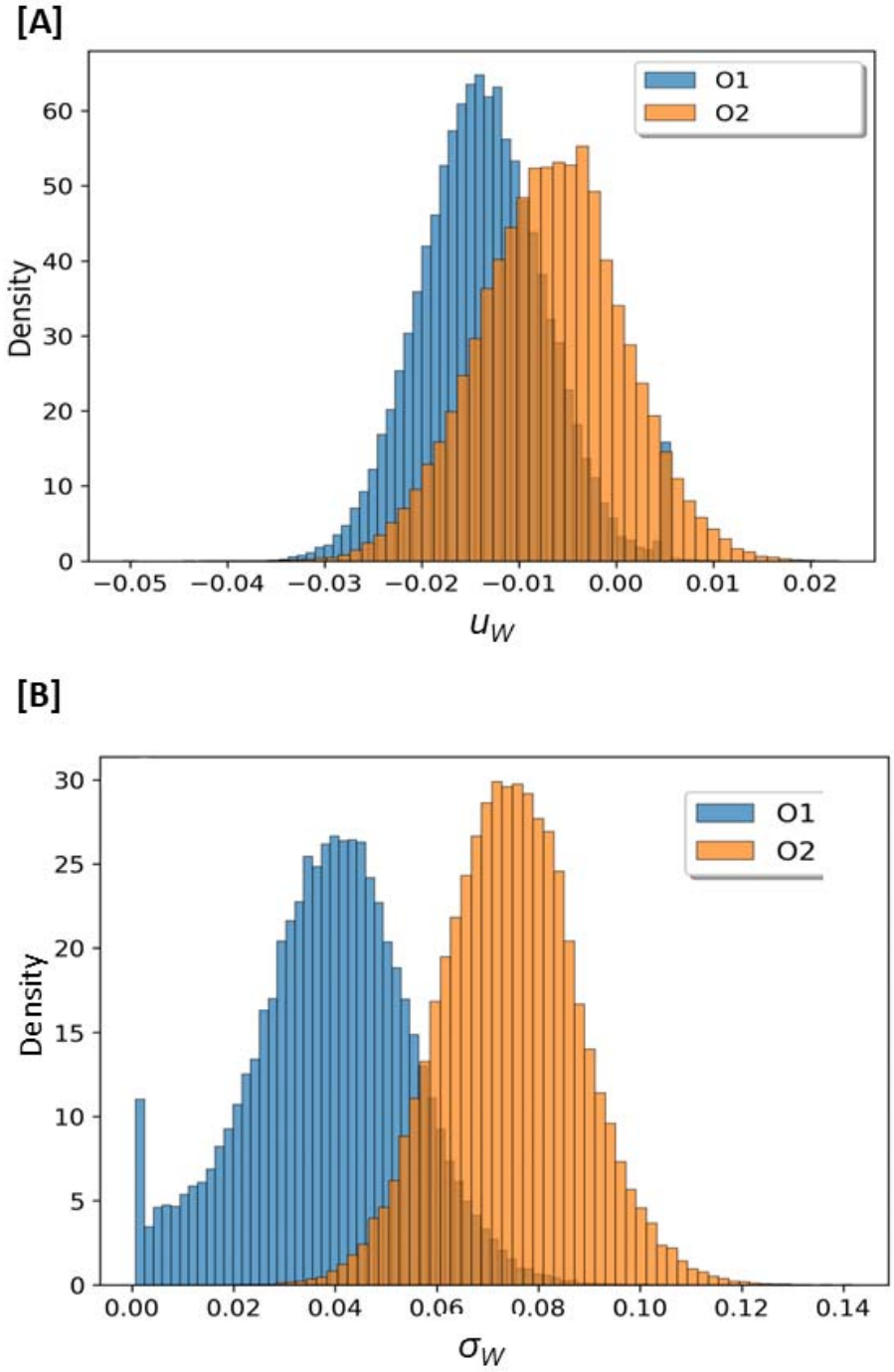


Figure 5. Distribution of posterior probabilities of mutational effects *u* on competitive fitness. (**A**) Mean *u* (**B**) Standard deviation σ_G.

It allows for a change in the mean mutation rate, provided that the mutational bias is in the direction of stronger selection.

SNV mutations are consistent with the predictions of the strong CoF model: the mean base substitution mutation rate μ_{SNV} did not change significantly over the course of ~300 generations of minimal selection, whereas the variance in μ_{SNV} increased significantly. Indels fulfill the predictions of the DB model but not the strong CoF. The indel mutation rate increased significantly over the course of the experiment; the point estimate of μ_{Indel} increased by about 2/3 from O1 to O2 and the number of false positive indels would have to be much greater than the estimated number to explain the observed increase in mutation rate. However, the data do not allow us to rule out the weak CoF, which posits that if a mutational bias exists, then the bias should be in the direction of stronger selection (Supplementary Figure 3). The trend for a negative association between competitive fitness and mutation rate (Supplementary Figure 6) is consistent with both the DB and the weak CoF, given the mutational bias for increased indel mutation rate.

The findings of this study reinforce and extend previous findings from *C. elegans* and other organisms, in which various types of physiological stress – exogenous and endogenous – increase the indel mutation rate, but have little if any effect on the base substitution mutation rate. Rajaei et al. (Rajaei, et al. 2021) performed an MA experiment with an N2-derived strain of *C. elegans* carrying a mutation (*Mev-1*) that causes increased cellular oxidative stress. In that experiment, the indel (and specifically, the insertion) rate increased by about 50% over wild-type, whereas the SNV mutation rate decreased slightly. Similarly, MA lines of *Drosophila melanogaster* initiated from strains engineered to carry large-effect deleterious mutations experience a higher rate of short deletions than their wild-type controls, whereas the SNV mutation rates do not differ (Sharp and Agrawal 2016). *Arabidopsis thaliana* MA lines propagated under benign and stressful conditions of salinity (Jiang, et al. 2014) and temperature (Belfield, et al. 2020) show large and highly significant increases in indel mutation rates

compared to much smaller increases in SNV rates. A similar increase in indel frequency was reported for the green alga *Chlamydomonas reinhardtii* MA lines maintained under salinity stress (Hasan, et al. 2022).

Our second-order MA experiment has three antecedents. First, Ávila, et al. (2006) initiated an O2MA experiment from a single high-fitness MA line of *D. melanogaster*. Their study predated economical whole-genome sequencing, but they found that the decline in fitness in the O2MA experiment was greater than in the O1MA experiment. They attributed the increased rate of decline in fitness to an increase in mutation rate, although they could not formally rule out greater mutational effects in the O2MA experiment. The subsequent findings of Sharp and Agrawal (2016) add credence to their conclusion. Second, Saxena, et al. (2019) did an O2MA experiment with *C. elegans*, designed to test the hypothesis that genomes that carry a high mutation load mutate faster than less-loaded genomes (Agrawal 2002; Shaw and Baer 2011). The indel rate (specifically the insertion rate) increased significantly in the O2MA lines; contrary to expectation, the O2MA indel rate was greater in the high-fitness (= less-loaded) lines. The SNV rate increased on average, but not significantly. Two of the ten O1MA families (n=5 O2MA sub-lines per O1MA family) evolved SNV O2 mutation rates that were significantly different from the ancestor; one family increased and the other decreased.

The third antecedent is the study of Liu and Zhang (2021), who brought the power of yeast genetics to bear on the question of mutation rate evolution. They first constructed a set of O1MA lines in which a mismatch repair gene (*MSH-2*) was deleted by CRISPR-Cas9 engineering to increase the mutation rate. They then constructed a set of O2MA lines by re-inserting the *MSH-2* gene into the O1MA lines and allowing mutations to accumulate further. They found (*i*) that the O2MA mutation rate increased on average, as predicted by the DB, but also (*ii*) 12/48 O2MA lines evolved a decreased mutation rate. They interpreted the high frequency of lines evolving decreased mutation rate as inconsistent with the DB hypothesis and consistent with the yeast mutation rate being maintained by apparent stabilizing selection (i.e.,

the weak CoF). Moreover, they identified a gene with antagonistic pleiotropic effects on cellular physiology (autophagy) and mutation rate, as predicted by the CoF. Their results, if not their conclusions, are in agreement with those of Saxena, et al. (2019) in *C. elegans*.

In 1995, Kondrashov (1995) famously posed the question "Why have we not died 100 times over?". Given the sizes of the genome (large) and effective population size (much smaller), the continued existence of humans seemed incompatible with the inferred genetic load (~100 lethal equivalents) if the effects of deleterious mutations are independent. The simplest explanation is that the effects of deleterious mutations are not independent, but rather they interact synergistically, i.e., epistasis is negative. However, it is not immediately obvious why deleterious mutations should interact synergistically on average. One possible explanation is that selection is "soft" (Wallace 1968), wherein population absolute fitness is established by resource availability and relative fitness results from competition among individuals for resources (Peck and Waxman 2000; Agrawal and Whitlock 2010). If so, it implies that epistasis is more likely to be detected under competitive conditions, as we report here.

This study provides no support for the possibility that deleterious mutations interact synergistically, and in fact the point estimate of the effect of second-order mutations (u_{O2}) is about half that of the first-order effect, suggestive of diminishing returns epistasis. However, we can confidently reject the two-stage model of mutational effects in favor of the one-stage model of constant mutational effects ($\Delta_{\mathrm{WAIC}} \approx 12$). There are obviously many caveats associated with that inference, including environmental context and statistical power. We note that our findings recapitulate those of Peters and Keightley (2000) and Katju, et al. (2015) who looked for epistatic effects of on non-competitive fitness in *C. elegans* and did not find them.

Conclusions. The results of this study are clear: there is an upward mutational bias for indels, whereas there is no such bias for SNV mutations, although the variance in μ_{SNV} increased over the course of the experiment. The discrepancy between evolution of the indel and SNV

mutational processes, in this study and others, is intriguing. One possibility is that the genomic mutational target for the control of the indel process is simply larger than that of SNVs, and that if selection was relaxed for long enough that the SNV mutation rate would begin to drift upward. We suspect this is probably true. However, it is worth considering the possibility that the indel and SNV mutational processes are subject to different patterns of selection, and that the DB or weak CoF model applies to indels, whereas the strong CoF model applies to SNVs.

What remains to be determined is the nature of the fitness function with respect to mutation rate. Liu and Zhang (2021) reported that mutation rate of their yeast O2MA lines is uncorrelated with growth rate. Visual inspection of their plot of growth rate against mutation rate (their Supplementary Figure 2E) reveals a weak positive relationship between mutation rate and growth rate, contrary to the prediction that lines with high mutation rates (mutators) should have lower than average fitness, as expected if mutation rate is under stabilizing selection in the presence of a mutational bias toward increased mutation rate. However, growth rate was measured in only one environment and both mutation rate and fitness in yeast are known to be context dependent (Liu and Zhang 2019), so it remains entirely possible that mutation rate in yeast is under global stabilizing selection. What Liu and Zhang clearly did show is (*i*) antimutator mutations are not trivially rare in yeast, but also (*ii*) on average, mutation rate drifts upward when selection is relaxed. We are aware of no case in which the average mutation rate has drifted downward.

The idea that epistasis among deleterious mutations should be synergistic on average has been as influential among evolutionary geneticists as the supporting evidence for it is weak (de Visser, et al. 2011). Beyond the obvious possibility that 50 years of inconclusive evidence is the result of underpowered experiments intended to quantify an inherently noisy trait, the intriguing possibility exists that epistasis itself is not only evolvable (Azevedo, et al. 2006; Desai, et al. 2007), but that the direction of epistasis favored by selection is dependent on idiosyncratic

properties of the population including effective population size and mutation rate (Gros, et al. 2009; Sydykova, et al. 2020).

Methods

Mutation Accumulation (MA) Experiment

We obtained a stock of the EG8072 strain of *C. elegans* (https://cgc.umn.edu/strain/EG8072) from the Caenorhabditis Genetics Center (CGC) at the University of Minnesota in September 2018. EG8072 is a derivative of the standard N2 strain, with green (GFP) and red (mCherry) fluorescent reporters stably integrated into the genome. We propagated the founder stock for seven generations by transfer of a single immature hermaphrodite at 4-day intervals, at which point the population is expected to have reached mutation-drift equilibrium ($6N_e$ generations; Lynch and Hill 1986). The population was expanded to large size (thousands of mixed-stage individuals) for 2-3 generations and cryopreserved.

In November 2018 we thawed an aliquot of the cryopreserved stock, cleaned it by bleaching (Stiernagle 2006) and initiated 148 replicate populations, each initiated from a single immature hermaphrodite. 48 replicates were expanded to large size (N ≈ a few thousand) for an additional 2-3 generations and cryopreserved in 20 replicate PCR plates. These are the G0 ancestral “pseudolines”. The remaining 100 populations – the first-order (O1) MA lines - were propagated by transfer of a single immature hermaphrodite at four-day intervals for an additional 600 days (Gmax=150 generations). MA lines were maintained under standard conditions: on 35 mm NGMA agar plates spotted with 50 µl of an overnight culture of *E. coli* OP50, kept in a Percival incubator set to a constant 20° C. Plates were kept on cafeteria trays, in 6 × 8 arrays, 48 plates per tray. Even numbered and odd numbered replicates were kept on different trays.

After the leading-generation worm was transferred, the plate it was taken from was moved to a different incubator set to 10° C; this is the (first) backup plate. Worms do not

reproduce (or reproduce only very slowly) at 10° C. If at the time of transfer a worm of the leading generation failed to reproduce or was not visibly gravid, it was replaced by a sibling taken from the previous generation plate. If there were offspring on the leading generation plate but no worm had reached at least the L2 stage, the plate was “held over” without transferring a worm. This protocol assures that generation number of a line is known with (near) certainty. Demographic data on backups and holdovers is given in Supplementary Table 6.

All lines were cryopreserved at 50 generation intervals. After 150 transfers (600 days, Gmax=150), each O1MA line was replicated into two second-order (O2) MA lines; the two O2MA lines derived from an O1MA line are an “O1MA family”. The O2MA lines were propagated as described for an additional ~150 generations (Gmax=306), at which time they were cryopreserved analogously to the G0 pseudolines as described above.

Genome Sequencing and Variant Calling

(i) <u>DNA Extraction and library preparation</u>. Cryopreserved tubes of worms were thawed and grown for several days on 60 mm NGMA agar plates containing Streptomycin and Nystatin and seeded with 110 µl of overnight culture of the HB101 strain of *E. coli*. When food was nearly exhausted, a small piece (~2 cm^2) of the agar plate was transferred onto a 100 mm NGMA plate and worms were allowed to grow for another 2-3 days. When food was nearly exhausted, worms were collected into 15 ml centrifuge tubes on ice and allowed to settle. Approximately 100 µl of settled worms were transferred into a 1.5 ml microcentrifuge tube and stored at -80°C until all samples were ready for DNA extraction, at which time samples were thawed, and genomic DNA was extracted with the DNeasy Kit (Qiagen) following the manufacturer's protocol. Extracted DNA was stored at -20°C prior to library preparation.

Frozen DNA samples were thawed and diluted with diH_2O to a concentration of 0.2 ng/µl. Tagmentation was performed per manufacturer’s instructions (Illumina, Nextera DNA Sample preparation kit, FC-121-1030). Following tagmentation, amplification was performed

using custom barcoded IDT primers. Following amplification, a 96-pooled well sample was generated. Once pooled, 170µL of sample was combined with 30µL of 6X loading dye and run on a 2% agarose gel. DNA segments of 300-500 bp were excised from gel. The QIAquick Gel Extraction Kit (QIAquick, 28704) was used to clean and elute DNA from the gel, following manufacturer's instructions. Pooled library samples were quantified using the Qubit HS kit (Qubit, Q33230) and sequenced to ~25X average depth of coverage with 150 bp paired-end Illumina sequencing on a NovaSeq 6000 by Novogene, Inc.

(ii) Genome Assembly. Raw sequence reads were filtered for quality (MQ≥3) and adapter sequence trimmed with fastp (Chen, et al. 2018). Reads were then aligned to the *C. elegans* N2 strain reference genome (WS292) using bowtie2(Langmead and Salzberg 2012). The resulting BAM files were sorted and filtered with Samtools (Li, et al. 2009) and bamTools (Barnett, et al. 2011) for downstream processing. Read group information was added with Picard (http://broadinstitute.github.io/picard). Duplicate reads were identified and removed using MarkDuplicates in GATK (McKenna, et al. 2010). Sequence quality was assessed with FastQC (Andrews 2010) and MultiQC (Ewels, et al. 2016) after trimming and alignment and prior to variant calling.

(iii) Variant Calling Pipeline. Variants were called with the HaplotypeCaller tool in GATK (McKenna, et al. 2010), run in GVCF mode. HaplotypeCaller identifies both single nucleotide variants (SNVs) and indels by local de novo assembly of haplotypes in regions with evidence of variation, improving accuracy by reconstructing the sequence context in those regions. GVCFs from all samples were merged with GenomicsDBImport, and joint genotyping was performed using GenotypeGVCFs. We used bcftools (Li 2011) to identify homozygous variants unique to a line, which were counted as putative mutations.

Variants that met the following three criteria were considered as putative mutations: (1) they were called homozygous; (2) they were present in one and only one O1MA family, and (3) the ancestor genotype is homozygous wild-type. If the variant is present in both O2MA sublines,

the mutation is inferred to have occurred in the first 150 generations and be a first-order (O1) mutation. If the variant is present in only one of the two O2MA sublines, it is inferred to have occurred subsequent to the split of the two O2MA sublines and be a second-order (O2) mutation. Comparisons of O1 and O2 mutation rates were restricted with the additional criteria (4) a site must be called in both members of an O1MA family and (5) the site must be called in at least 90% of O1MA families.

One O2MA line from O1MA family 9 was not sequenced. The identities of two O2MA lines appear to have been swapped, one from O1MA family 34 and one from family 46. One O1MA line (family 26) evolved an indel mutator phenotype, as evidenced by the >5X greater indel rates of the two O2MA sublines than of the O2MA line with the next highest indel rate (Supplementary Table 1). Except where noted, those families were omitted from further analysis.

In the O2MA lines, over 97% of the ~100.2 Mb genome was covered at the liberal 3X coverage criterion (mean= 97.6%, median = 97.8%; Supplementary Table 1). At 3X coverage, the mean and median fractions of the genome included in the O1MA analysis are 97.3% and 97.5% respectively. At 10X coverage the O2MA mean and median are 93.5% and 95.8%; the O1MA mean and median are 89.8% and 93.2%.

Assessment of falsely inferred variants

(*i*) <u>False positives</u>. EG8072 is derived from the N2 strain, which is the source of the *C. elegans* reference genome (Ichikawa, et al. 2025). Any difference between our EG8072 ancestor and the reference genome is either (1) a mutation that fixed subsequent to the divergence of EG8072 and the N2 reference strain, (2) a false negative in the N2 reference, or (3) FP in our EG8072 ancestor. In the first two cases, all MA lines derived from the EG8072 ancestor should also have the variant allele. However, if the variant is FP in our EG8072 ancestor, the derived MA lines should all have the reference allele. The EG8072 G0 ancestor was sequenced from

three independent Illumina libraries. We estimated the frequency of FPs from each of the three samples of the ancestor.

At the liberal (3X) coverage criterion we identified an average of 1662 fixed variants between our G0 ancestor and the reference genome (1676, 1647, and 1664 in the three replicates, respectively; Supplementary Table 3), of which approximately 42% were single nucleotide variants (SNVs) and 58% were insertions or deletions (indels). In the three replicates we identified three (1 deletion, 2 SNVs), three (3 SNVs) and one (1 SNV) FPs respectively. One of the SNVs (a G:C→C:G transversion) was common to all three replicates; the others were only found in one replicate. Of the seven FPs detected at the liberal 3X criterion, six were also found at the stringent 10X criterion. By way of comparison, the same analysis with a different N2 ancestor found one FP (an SNV) out of 1439 variants with 68 MA lines at the 3X criterion (Rajaei, et al. 2021).

(*ii*) <u>Failure to recall (FtR) and false negatives (FN)</u>. A mutation can fail to be detected ("failure to recall", FtR) for two reasons. First, a site with a true mutant may not be called (i.e., missing data). Second, a true mutant allele may be mistakenly identified as an ancestral allele; that is a false negative (FN). FNs are a subset of FtR and cause the mutation rate to be underestimated. FtR due to missing data may or may not affect estimation of the mutation rate, depending on whether sites with true mutations are more or less likely to be called than sites without mutations. All else equal, sites with true mutations are less likely to be called than sites without mutations. The reason is that if a sequence with *n* or more mismatches exceeds the cutoff for mapping, a sequence with a mutation has one more mismatch than the same sequence without the mutation. All else also equal, FNs are equally likely to be true O1 mutations or true O2 mutations. Since there are twice as many opportunities for O2 as O1 mutations (Supplementary Figure 3), FNs introduce a downward bias in the O2 mutation rate relative to the O1 mutation rate.

To assess our ability to recall mutations, we introduced simulated “dummy” mutations at random into the reference genome and analyzed the data using our standard variant-calling pipeline, but with the simulated genome as the reference (Rajaei, et al. 2021). The changed site in the reference should appear as a variant in our sequenced sample. For example, suppose the true reference allele at a position is an A and we change it to a C. The site should be called variant (A/A in this example) in our data. If the site is either not covered in the data or is called anything other than a variant homozygote (A/A), we have failed to recall the mutation. If the site is covered in the data but not called a variant homozygote, that is FN.

We constructed a pseudo-reference genome by inserting 1200 single-nucleotide variants (SNVs) and 12,000 indels (1–20 bp, size chosen at random) into the *C. elegans* N2 (WS292) reference genome (Supplementary Table 7). First, we randomly inserted “dummy” SNVs and indels into the reference genome in random positions using the R packages GenomicRanges, IRange, and Biostrings and compiled them into a VCF file. The VCF was then sorted and indexed with samtools, and the pseudo-reference genome was produced with GATK FastaAlternateReferenceMaker. The genome assembly and variant calling pipelines were rerun with all the MA lines and the G0 ancestor using this pseudo-reference genome. Line-specific simulated FtR and FN rates and counts are given in Supplementary Table 8.

To validate the consistency of the procedure, the original variants (i.e., putative mutations) were re-mapped against the pseudo-reference genome. All sites identified as variant in the original analysis were also called as variant against the pseudo-reference. We repeated the simulation twice independently and obtained consistent results across runs (Supplementary Table 9).

At the liberal 3X criterion, the average FtR rate was ~2.2% for SNVs and ~2.5% for indels. The FN rates were 0.21% and 0.53% for SNVs and indels. At the stringent 10X criterion the failure to recall rates were 7.8% and 7.7% for SNVs and indels respectively; the FN rates

were 0.084% and 0.12%. All analyses are based on the 3X coverage data; 10X data are reported in the Supplementary Tables.

Competitive Fitness Assay

(*i*) Focal worms. 50 G150 O1MA lines and 20 G0 pseudolines (PS lines) were chosen at random for inclusion in assay block 1. Block 2 included one G300 O2MA replicate from each of those same 50 lines and the same 20 PS lines; Block 3 included the other O2MA replicate, along with the same 20 PS lines. At the beginning of each assay (*Day 1*) a cryopreserved aliquot from each line (MA and PS) was thawed, and ~150 µl pipetted onto a 60 mm NGMA plate seeded with the 100 µl of an overnight culture of the OP50 strain of *E. coli* (the thaw plate; Figure 2B) Two days later (*Day 3*), seven immature worms (L3/L4 stage) from each line were picked singly to a 35 mm NGMA plate seeded with 50 µl of OP50. These plates are the parental (P) generation. Each P plate was identified with a random number and arranged in numerical order on cafeteria trays, 48 plates/tray, and placed in a Percival incubator maintained at 20˚C and ambient humidity. Thaw plates were incubated at 10˚C to prevent further reproduction. Four days later (*Day 7*), a single immature (L3/L4) hermaphrodite was picked to a new plate and maintained as before. These plates are the F1 generation. If a P generation worm failed to reproduce, it was replaced in the assay with a sibling from the thaw plate. Four days later (*Day 11*) the process was repeated for the F2 generation. If an F1 worm failed to reproduce, the replicate was removed from the assay and treated as missing data. Three days later (*Day 14*) a single L1 stage worm was picked from each F2 plate onto the assay plate, along with an L1 competitor strain worm.

(*ii*) Competitor worms. On *Day 1* of the assay, an aliquot of N2 strain worms was thawed and pipetted onto a seeded 35 mm plate and allowed to reproduce. When the food was exhausted (around *Day 5*), a ~2 cm chunk was transferred to a new 100 mm NGMA plate seeded with 400 µl of OP50 and allowed to reproduce. When food was exhausted (around *Day 9*), a chunk was

transferred to a new plate. On *Day 13*, worms were washed from the plate and bleached to synchronize the population. Bleached embryos were pipetted in equal aliquots onto four 100 mm NGMA plates supplemented with streptomycin and the antifungal agent nystatin and seeded with 400 µl of the HB101 strain of *E. coli*.

(*iii*) Assay plates. A single L1 focal worm (MA or G0 pseudoline) and a single L1 competitor worm were picked to a 35 mm NGMA plate supplemented with streptomycin and nystatin and seeded with 50 µl of HB101. Assay plates were identified by random number, placed in random order on trays (48 plates/tray) and incubated at 20°C for eight days (*Day 21*; approximately two overlapping generations), at which point nearly all plates were depleted of bacterial food. Worms were washed from the assay plate in 2 ml of M9 buffer and transferred to a 2 ml deep-well 96-well plate. The plate was centrifuged for 1 min at 1,000 g, and the supernatant was aspirated with an 8-channel strip aspirator set to leave 100 µl in each well. The worms were washed once more by adding 1.4 ml of M9 buffer, centrifuging, and aspirating as before, followed by the addition of 1.4 ml of M9. Worms were mixed by gently pipetting with a 12-channel pipette and a 110 µl sample was transferred into a well in a 96-well, clear-bottom tissue culture plate. 5 mM levamisole was added to each well to immobilize worms prior to imaging.

(iv) Imaging. Bright-field and GFP fluorescence images (470±20 nm excitation, 525±25 nm emission) were captured at 20x or 30x magnification for each well using an automated epifluorescence microscope (IX-70, Olympus, Pittsburgh, PA, USA) fitted with a CCD camera (Retiga-2000R, QImaging, Surrey, BC, Canada), XYZ stage and focus motors (Prior Scientific, Cambridge, UK), and controlled by Image Pro-plus software (Media Cybernetics, Rockville, MD, USA). 96-well plates were imaged first under bright field (BF) and then under fluorescence (GFP). Images of the same well (BF and GFP) were paired and renamed after checking by eye that the images matched.

(v) Counting. Worms were counted from stacked images using ImageJ software. We summarize the protocol here; a detailed tutorial of the ImageJ analysis is included as Supplementary Appendix 1.

Step 1. Paired images were stacked using the "Stacks" option under the "Image" tab.

Step 2. Images were merged and colorized using the "Merge Channels" tool under the "Image" tab, with fluorescent worms colored differently from non-fluorescent ones.

Step 3. For best effect, brightness and contrast of GFP images were adjusted by setting "minimum" and "maximum" light low and "contrast" high under the "Image" tab.

Step 4. If an image was too crowded with worms, the ROI selection tool in the "Analyze" tab was used to crop a sample portion of the stacked images to count. The stacked images were synchronized using Analyze →Tools →Synchronize windows → Synchronize All commands.

Step 5. Worms were counted by eye using the "Multipoint" tool to mark individual worms, with different marking colors for fluorescent and non-fluorescent worms.

Data Analysis

i) Mutation Rate. We calculated the per-generation mutation rate, μ_i, for each MA line *i* as $\mu_i = m_i/n_it_i$, where *m* is the number of mutations, *n* is the number of callable sites, and *t* is the number of generations of MA (Denver, et al. 2009). First-order and second-order mutation rates were calculated separately.

Our first question of interest is: does the mutation rate change over time? To test that hypothesis, we can estimate the mutation rates of O1MA lines (μ_{O1}) and O2MA lines (μ_{O2}), calculate the difference $\Delta\mu = \mu_{O2} - \mu_{O1}$, and test the hypothesis that the average difference is significantly different from zero. There is a complication, however, because false inferences affect estimates of μ_{O1} and μ_{O2} in different ways (Supplementary Figure 3).

To proceed, we employed a parametric bootstrap approach. We assume that mutation can be modeled as a Poisson process. We further assume that there is a uniform mutation rate $\mu = \mu_{O1} = \mu_{O2}$, i.e., the mutation rate does not change over the course of the experiment. Because FPs necessarily inflate μ_{O2} relative to μ_{O1} and FNs do the opposite on average, ignoring FNs leads to a conservative test of the prior hypothesis that $\mu_{O2} > \mu_{O1}$. To begin, we set μ= to the point estimate of μ_{O1}. At each generation, an MA line accumulates mutations in its genome drawn from a Poisson distribution with parameter $\lambda_m = \mu L$, where μ is the per-generation mutation rate and *L* is the size of the *C. elegans* genome. The total number of mutations in line *i*, $m_i = \lambda_m t_i$, where t_i is the number of generations of MA for line *i*. At Gmax=150, the O1MA line is split into the two O2MA sublines, and the mutational process is continued for the remaining Gmax=150 generations. Each mutation was accepted ("called") or rejected with binomial probability equal to the proportion of the genome covered in the line (Supplementary Table 1). The number of generations for each MA line was held constant to the observed number for the line. Upon reaching Gmax=300, *n* FPs were added to each O2MA line, drawn from a Poisson distribution with parameter λ_{FP}= the expected number of FPs per line. We next calculate μ_{O1} and μ_{O2} from the simulated data and subtract μ_{O1} from μ_{O2} to find $\Delta\mu$. This procedure was repeated 10,000 times to generate a null distribution of $\Delta\mu$. The value of λ_{FP} for which the observed value of $\Delta\mu$ exceeds 95% of simulated values is the maximum number of FPs per O2MA line that is consistent (at p>0.05) with $\mu_{O2} = \mu_{O1}$, i.e., $\Delta\mu > 0$ cannot be explained by FPs if $\mu_{O2} = \mu_{O1}$.

ii) Variance in mutation rate. Also of interest is the question: does the variance in mutation rate (V_μ) change over time? We address this question in two ways. First, from the same simulations described in the preceding section. For each simulation run we calculate the variance among O1MA lines ($V_{\mu O1}$) and O2MA lines ($V_{\mu O2}$) to generate the relevant null distributions. We then

compare the observed variances to the null distributions to test the hypothesis that there is more variance in mutation rate than expected if the mutation rate is uniform(Saxena, et al. 2019).

Second, the design of the experiment permits a simple test of the heritability of mutation rate over the time frame of the first-order MA (≈ 150 generations). The phenotypic correlation between relatives $\rho_{x,y} = r_{x,y}h^2$, where $r_{x,y}$ is the coefficient of relatedness between individuals *x* and *y* and h^2 is the narrow-sense heritability (Gillespie 2004, Equation 6.4). The two O2MA sublines of an O1MA family are identical twins at the time of "birth" (Gmax=150), and $r_{x,y}$ = 1. Thus, the heritability of mutation rate is simply the correlation between the mutation rates of the O2MA sublines.

iii) Fitness effects. Competitive fitness *W* is defined as the logarithm of the ratio of the frequency of focal strain worms (*p*) to competitor strain worms (1-*p*) on an assay plate, i.e., *W* = log[*p*/(1-*p*)]. To quantify how mutations contribute to competitive fitness, we applied a fully Bayesian mixed-effects modeling framework (Stearns, et al. 2025). We compared two Bayesian models: a one-stage ("uniform effects") model that consolidates all mutations into a single genetic effect structure using a unified genetic covariance matrix, and a two-stage model that explicitly partitions mutations into first-order (O1) and second-order (O2) effects, each with its own genetic parameters and separate covariance matrices. Each value of *W* for observation *i* is modeled as the sum of a global intercept (*c*), a block-specific deviation ($b_{\text{block}[i]}$), genetic effects (g_i), and a residual error (ε_i):

$$W_i = c + b_{block[i]} + g_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_e^2).$$

The block effects (b_j) account for systematic batch-to-batch variability arising from separate experimental runs, modeled as independent deviations from a normal distribution:

$$b_j \sim N(0, \sigma_b^2).$$

The genetic effects (g_i) capture heritable variation in *W* explained by genetic similarity among strains. Following the Central Limit Theorem, we assume that the cumulative effect of multiple

mutations can be approximated by a normal distribution. For each individual *i*, the genetic effect is decomposed into two additive components:

$$g_i = g_{1,i} + g_{2,i},$$

where $g_{1,i}$ represents the genetic effect from first-order mutations (accumulated from generations 0 to 150) and $g_{2,i}$ represents the genetic effect from second-order mutations (accumulated from generations 150 to 300). Each genetic effect component is assumed to follow a normal distribution with mean and variance proportional to the number of mutations:

$$g_{1,i} \sim N(n_{1,i}\mu_1, n_{1,i}\sigma_1^2) \quad \textit{and} \quad g_{2,i} \sim N(n_{2,i}\mu_2, n_{2,i}\sigma_2^2),$$

where $n_{1,i}$ and $n_{2,i}$ represent the number of mutations for individual *i* in the first-order and second-order experiments, respectively. The parameters μ_1 and μ_2 represent the mean fitness effect per mutation, while σ_1^2 and σ_2^2 represent the variance in fitness effects per mutation. In the one-stage ("uniform effects") model, we model the total genetic effect for each individual using a single genetic covariance matrix (*G*) that combines both O1 and O2 mutations:

$$g_i \sim N(n_{total,i}\mu_g, n_{total,i}\sigma_g^2), \quad n_{total,i} = n_{1,i} + n_{2,i},$$

where μ_g and σ_g represent the overall genetic mean and scale parameters, and the covariance structure is captured by a unified *G* matrix incorporating all mutations. In the two-stage model, we model the genetic effects separately using independent covariance matrices (G_1 for O1 mutations and G_2 for O2 mutations):

$$g_{1,i} \sim N(n_{1,i}\mu_{g1}, n_{1,i}\sigma_{g1}^2) \quad \textit{and} \quad g_{2,i} \sim N(n_{2,i}\mu_{g2}, n_{2,i}\sigma_{g2}^2),$$

where μ_{g1}, μ_{g2}, σ_{g1}^2, and σ_{g2}^2 represent the separate genetic parameters for first-order and second-order mutations, and the covariance structure is partitioned between G_1 and G_2 matrices.

We implemented both models using PyMC (Salvatier, et al. 2016) with weakly informative priors to regularize estimation. Intercepts and genetic means were assigned *N*(0,10) priors, while standard deviations received HalfNormal(5) priors. Model parameters were estimated using the No-U-Turn Sampler (NUTS) with 2,000 tuning and 20,000 posterior draws.

To account for potential measurement errors in mutation counts, we implemented a bootstrapping approach that simulates both false positives and false negatives. False positives, which only affect O2 mutations, were modeled using a Poisson distribution with expectation $\lambda = 3$. False negatives, which in this context include all mutations missed due to failure to recall, were also modeled using a Poisson distribution with rate

$$\lambda_{missed} = \mu_{mutation} \times n_{generations} \times genome_length \times failure_rate,$$

where $\mu_{mutation}$ represents the per-generation mutation rate (2.13×10^{-9} for O1, 2.47×10^{-9} for O2), $n_{generations} = 150$, genome_length = 10^8 bp (the approximate size of the *C. elegans* genome), and failure_rate = 0.025. For each bootstrap iteration, we resampled mutation counts by subtracting false positives from observed O2 counts, adding missed mutations to both O1 and O2 counts, and fitting the Bayesian model with the resampled mutation counts. This approach allows us to propagate uncertainty in mutation counts through to our parameter estimates while maintaining the mathematical structure of our genetic effects model.

Data Availability

Raw sequence data are deposited in the NCBI Sequence Read Archive BioProject #PRJNA1089216. All data in the Supplementary Tables are also deposited in Figshare at https://doi.org/10.6084/m9.figshare.30197935. Code and unprocessed image files are available at https://github.com/Baer-Group/O2MAno2.

Acknowledgments

David Julian generously provided access to his microscope for the fitness assays, and Tim Crombie assisted with microscope software. Ezgi Karabulut, Eileen Kelly, and Hope Kerr assisted with counting worms. Support was provided by National Institutes of Health awards GM127433 (CFB, VK, and ECA) and GM154908 (JZ).

Figure Legends

Figure 1. Theoretical models of mutation rate evolution. (**A**) Cost of Fidelity. (**B**) Drift Barrier.

Figure 2. (**A**) Schematic depiction of the MA experiment. Short black lines depict ancestral G0 "pseudolines" (PS), blue lines depict O1MA lines, orange lines depict O2MA lines. (**B**) Schematic depiction of the competitive fitness assay. Label "LX" refers to "Line X" (MA or G0), label LX.1,…LX.7 refers to replicates 1-7 of a line in the assay. See Methods for details.

Figure 3. Distribution of simulated values of $\Delta\mu=\mu_{O2}-\mu_{O1}$ with increasing number of false positives. Red dashed line is the observed value. "FP cutoff" is the minimum number of false positives necessary for the data to be consistent with the hypothesis $\Delta\mu=0$. (**A**) SNVs (**B**) Indels.

Figure 4. Type-specific mutation rates and spectra. (**A**) SNV mutation rates (**B**) SNV frequency spectrum (**C**) Indel mutation rates (**D**) Indel frequency spectrum.

Figure 5. Distribution of posterior probabilities of mutational effects u on competitive fitness. (**A**) Mean u (**B**) Standard deviation σ_G.

Supplementary Figure 1. Cost of Fidelity with mutational bias. Curves depict the relationship between fitness (W) and mutation rate (μ). Blue horizontal arrows represent the input of mutational variance for mutation rate; red vertical arrows represent the strength of stabilizing selection against (anti)mutator variants; $w(\mu) = \exp\left(-\frac{\mu^2}{2\sigma^2}\right)$. Thickness of the arrows depicts the magnitude of the effect.

Supplementary Figure 2. Distribution of mutations along chromosomes. Top panel, SNVs; bottom panel, indels.

Supplementary Figure 3. Schematic depiction of the effects of false inferences. Stem branch is the O1MA line, bifurcating terminal branches are the two O2MA lines derived from the O1MA

line. Mutations are shown represented as arrows on the lineage. True sequence inferences are shown in blue text at sequence position 6; false inferences are shown in red text.

Supplementary Figure 4. Simulated distributions of variance in line mean mutation rates under the assumption of uniform mutation rate and the observed average number of false positives (FP) per line. Vertical dashed lines show the observed values. (**A**) SNVs, 2 FPs/line (**B**) Indels, 0.33 FPs/line.

Supplementary Figure 5. Distribution of line mean competitive fitness *W* by treatment (MA vs G0) and assay block. (**A**) G0 pseudolines (**B**) G150 MA lines (**C**) G300 MA lines (**D**) Comparison across assay blocks.

Supplementary Figure 6. Plots of line mean fitness against mutation rate. (**A-C**) All data; (**E-G**) generation 300 data only.

References

Agrawal AF. 2002. Genetic loads under fitness-dependent mutation rates. Journal of Evolutionary Biology 15:1004–1010.

Agrawal AF, Whitlock MC. 2010. Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? Trends in Ecology & Evolution 25:450–458.

FastQC: a quality control tool for high throughput sequence data [Internet]. 2010. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ávila V, Chavarrias D, Sanchez E, Manrique A, Lopez-Fanjul C, Garcia-Dorado A. 2006. Increase of the spontaneous mutation rate in a long-term experiment with *Drosophila melanogaster*. Genetics 173:267–277.

Azevedo RBR, Lohaus R, Srinivasan S, Dang KK, Burch CL. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature 440:87–90.

Baer CF, Shaw F, Steding C, Baumgartner M, Hawkins A, Houppert A, Mason N, Reed M, Simonelic K, Woodard W, et al. 2005. Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. Proceedings of the National Academy of Sciences of the United States of America 102:5785–5790.

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27:1691–1692.

Barton NH. 1995. A general model for the evolution of recombination. Genetical Research 65:123–144.

Belfield EJ, Brown C, Ding ZJ, Chapman L, Luo M, Hinde E, van Es SW, Johnson S, Ning Y, Zheng SJ, et al. 2020. Thermal stress accelerates *Arabidopsis thaliana* mutation rate. Genome Research.

Charlesworth B. 1996. The good fairy godmother of evolutionary genetics. Current biology : CB 6:220.

Chen SF, Zhou YQ, Chen YR, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:884–890.

de Visser J, Cooper T, Elena S. 2011. The causes of epistasis. Proceedings of the Royal Society B-Biological Sciences 278:3617–3624.

Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. Proceedings of the National Academy of Sciences of the United States of America 106:16310–16314.

Desai M, Weissman D, Feldman M. 2007. Evolution can favor antagonistic epistasis. Genetics 177:1001–1010.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610–618.

Fisher RA. 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon Press.

Gillespie JH. 2004. Population Genetics: A Concise Guide. Baltimore, MD: Johns Hopkins University Press.

Gros P, Le Nagard H, Tenaillon O. 2009. The Evolution of Epistasis and Its Links With Genetic Robustness, Complexity and Drift in a Phenotypic Model of Adaptation. Genetics 182:277–293.

Hasan A, Lachapelle J, El-Shawa S, Potjewyd R, Ford S, Ness R. 2022. Salt stress alters the spectrum of de novo mutation available to selection during experimental adaptation of Chlamydomonas reinhardtii. Evolution 76:2450–2463.

Ichikawa K, Shoura M, Artiles K, Jeong D, Owa C, Kobayashi H, Suzuki Y, Kanamori M, Toyoshima Y, Iino Y, et al. 2025. CGC1, a new reference genome for *Caenorhabditis elegans*. Genome Research 35:1902–1918.

Jiang CF, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP. 2014. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. Genome Research 24:1821–1829.

Katju V, Packard LB, Bu L, Keightley PD, Bergthorsson U. 2015. Fitness decline in spontaneous mutation accumulation lines of Caenorhabditis elegans with varying effective population sizes. Evolution 69:104–116.

Keightley PD, Caballero A. 1997. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the United States of America 94:3823–3827.

Kimura M. 1967. On evolutionary adjustment of spontaneous mutation rates. Genetical Research 9:23–27.

Kondrashov AS. 1995. Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? Journal of Theoretical Biology 175:583–594.

Konrad A, Brady MJ, Bergthorsson U, Katju V. 2019. Mutational landscape of spontaneous base substitutions and small indels in experimental *Caenorhabditis elegans* populations of differing size. Genetics 212:837–854.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–U354.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data P. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liu HX, Zhang JZ. 2021. The rate and molecular spectrum of mutation are selectively maintained in yeast. Nature Communications 12.

Liu HX, Zhang JZ. 2019. Yeast spontaneous mutation rate and spectrum vary with environment. Current Biology 29:1584–1591.

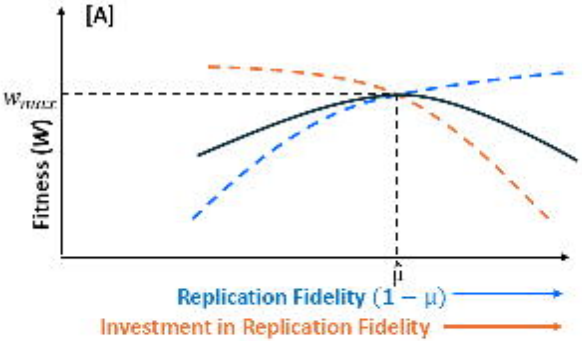
Lynch M. 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. Genetics 180:933–943.

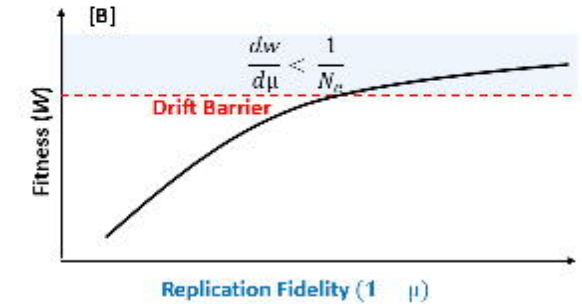
Lynch M. 2011. The lower bound to the evolution of mutation rates. Genome Biology and Evolution 3:1107–1118.

Lynch M, Hill WG. 1986. Phenotypic evolution by neutral mutation. Evolution 40:915–935.

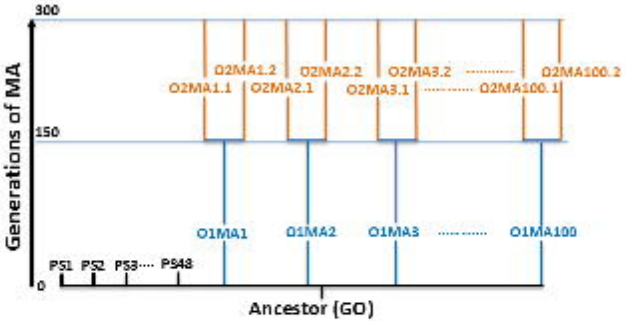
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce

framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–1303.
Otto SP. 2009. The evolutionary enigma of sex. American Naturalist 174:S1–S14.
Peck J, Waxman D. 2000. Mutation and sex in a competitive world. Nature 406:399–404.
Peters A, Keightley P. 2000. A test for epistasis among induced mutations in Caenorhabditis elegans. Genetics 156:1635–1647.
Rajaei M, Saxena AS, Johnson LM, Snyder MC, Crombie TA, Tanny RE, Andersen EC, Joyner-Matos J, Baer CF. 2021. Mutability of mononucleotide repeats, not oxidative stress, explains the discrepancy between laboratory-accumulated mutations and the natural allele-frequency spectrum in *C. elegans*. Genome Research 31:1602–1613.
Salvatier J, Wiecki TV, Fonnesbeck C. 2016. Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2:e55.
Saxena AS, Salomon MP, Matsuba C, Yeh SD, Baer CF. 2019. Evolution of the mutational process under relaxed selection in *Caenorhabditis elegans*. Molecular Biology and Evolution 36:239–251.
Sharp NP, Agrawal AF. 2016. Low genetic quality alters key dimensions of the mutational spectrum. PLoS Biology 14.
Shaw FH, Baer CF. 2011. Evolutionary consequences of fitness-dependent mutation rate in finite populations. Journal of Evolutionary Biology 24:1677–1684.
Stearns F, Zhou J, Fenster C. 2025. Scaling the fitness effects of mutations with respect to differentially adapted Arabidopsis thaliana accessions under natural conditions. Evolution 79:951–961.
Stiernagle T. 2006. Maintenance of *C. elegans*. In: Community T*Ce*R, editor. Wormbook.
Sturtevant AH. 1937. Essays on evolution. 1. On the effects of selection on mutation rate. Quarterly Review of Biology 12:467–477.
Sydykova D, LaBar T, Adami C, Wilke C. 2020. Moderate Amounts of Epistasis are Not Evolutionarily Stable in Small Populations. Journal of Molecular Evolution 88:435–444.
Vassilieva LL, Lynch M. 1999. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. Genetics 151:119–129.
Vehtari A, Gelman A, Gabry J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27:1413–1432.
Wallace B. 1968. Polymorphism, population size, and genetic load. In: Lewontin RC, editor. Population Biology and Evolution. Syracuse, NY: Syracuse University Press. p. 87–108.

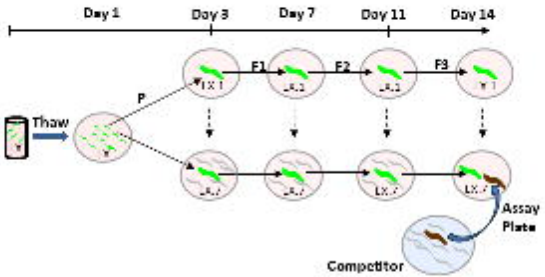

[A]
w_{max}
Fitness (W)
μ
Replication Fidelity (1 − μ)
Investment in Replication Fidelity


[B]
$\frac{dw}{d\mu} < \frac{1}{N_e}$
Drift Barrier
Fitness (W)
Replication Fidelity (1 μ)

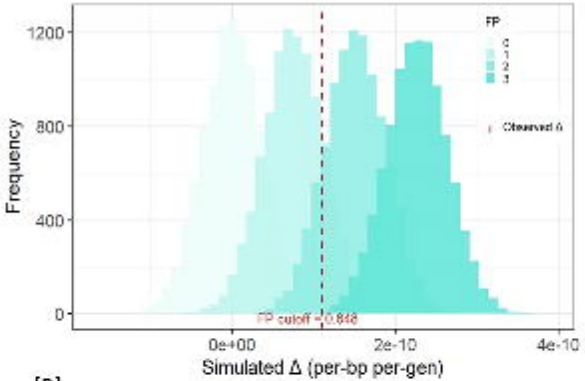
[A]


300
150
0
Generations of MA
O2MA1.2
O2MA2.2
O2MA3.2
O2MA100.2
O2MA1.1
O2MA2.1
O2MA3.1
O2MA100.1
O1MA1
O1MA2
O1MA3
O1MA100
PS1
PS2
PS3
PS48
Ancestor (G0)

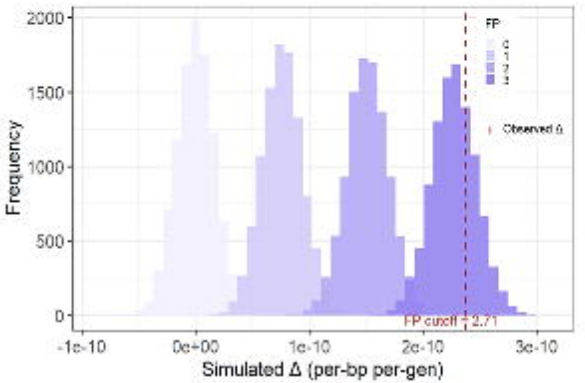
[B]

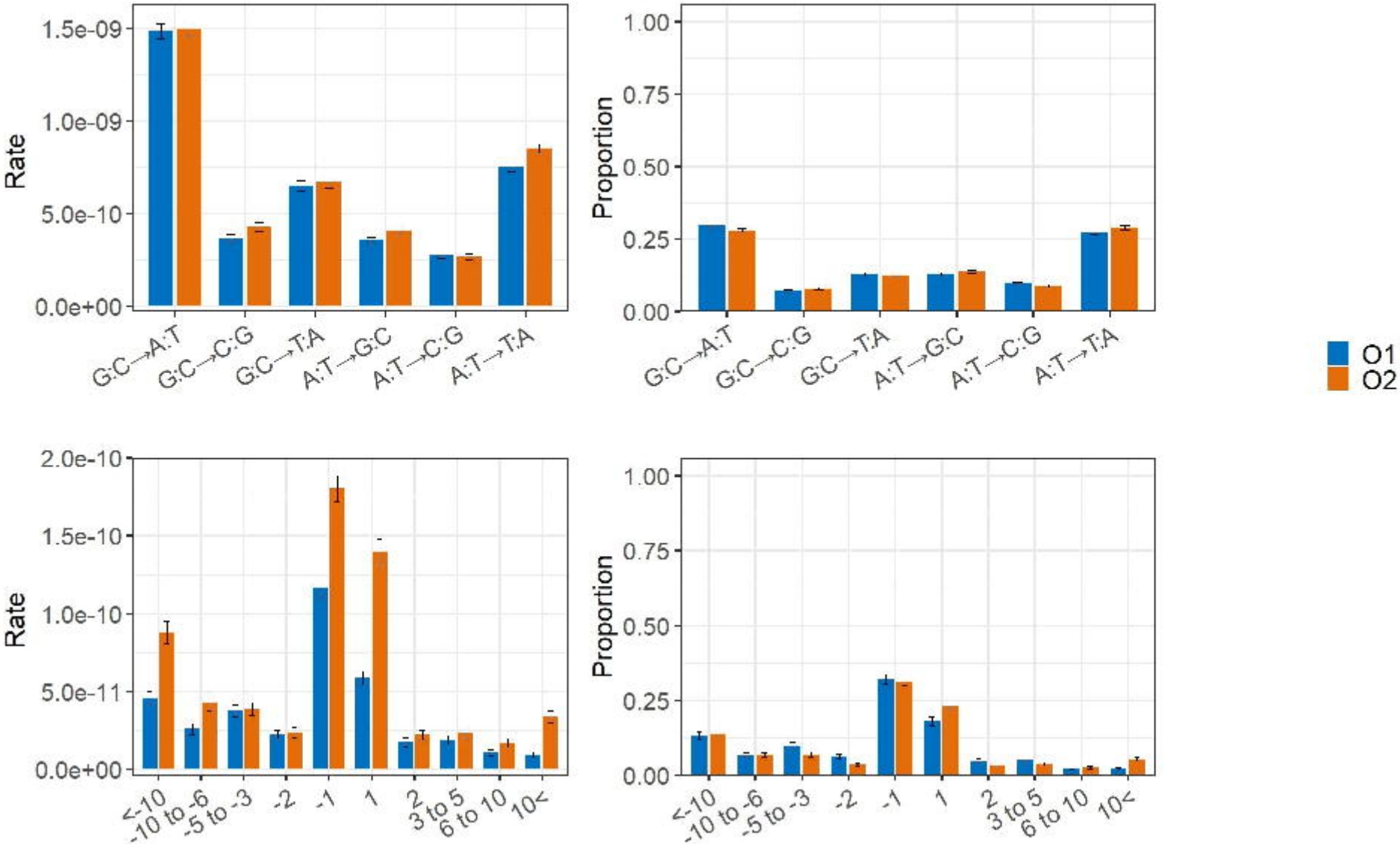

Day 1
Day 3
Day 7
Day 11
Day 14
F1
F2
F3
P
Thaw
Assay Plate
Competitor

[A]

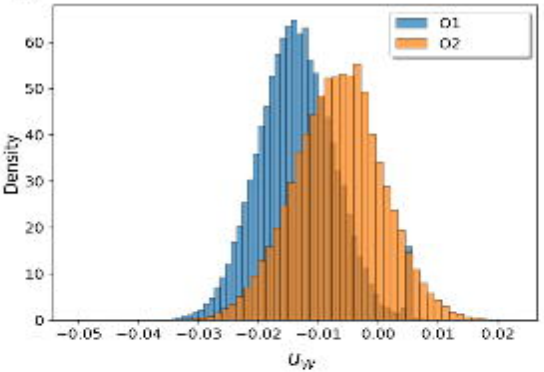

FP
0
1
2
3
Observed Δ
Frequency
1200
800
400
0
FP cutoff = 0.848
0e+00
2e-10
4e-10
Simulated Δ (per-bp per-gen)

[B]


FP
0
1
2
3
Observed Δ
Frequency
2000
1500
1000
500
0
FP cutoff = 2.71
-1e-10
0e+00
1e-10
2e-10
3e-10
Simulated Δ (per-bp per-gen)


Rate
1.5e-09
1.0e-09
5.0e-10
0.0e+00
G:C→A:T
G:C→C:G
G:C→T:A
A:T→G:C
A:T→C:G
A:T→T:A
Proportion
1.00
0.75
0.50
0.25
0.00
O1
O2
2.0e-10
1.5e-10
1.0e-10
5.0e-11
<-10
-10 to -6
-5 to -3
-2
-1
1
2
3 to 5
6 to 10
10<

[A]


O1
O2
Density
0
10
20
30
40
50
60
−0.05
−0.04
−0.03
−0.02
−0.01
0.00
0.01
0.02
$u_{/N}$

[B]

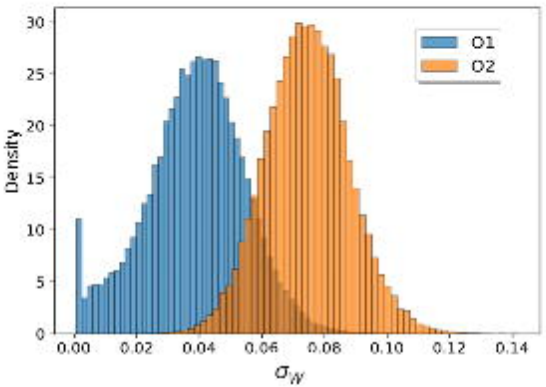

O1
O2
Density
0
5
10
15
20
25
30
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
$\sigma_{/N}$

Table 1. Summary statistics (SEM in parentheses). N_{Lines} is the number of MA lines, $\bar{t}$ is the average number of generations of MA (top row) and number of days of MA (bottom row). Values of $\bar{\mu}$ are mutation rates ($\times 10^9$) calculated from the observed number of mutations without correction for false inferences. Values of $\hat{\mu}$ are estimates of mutation rates ($\times 10^9$) corrected for false positives (FP) and mutations that were failed to recall (FtR). We assume the average number of FPs (2 SNVs and 0.33 indels) per genome and the average proportion of simulated mutations that were FtR at 3X coverage (2.2% SNVs, 2.5% indels). Estimated mutation rates ($\hat{\mu}$) are calculated as $\frac{n-\overline{FP}+\overline{FtR}}{L\times t}$, where n is the observed number of variants and L (100,286,401 bp) is the size of the *C. elegans* N2 reference genome.

	N_{Lines}	*%Genome*	$\bar{t}$	$\bar{\mu}_{SNV}$	$\bar{\mu}_{DEL}$	$\bar{\mu}_{INS}$	$\bar{\mu}_{TOTAL}$	$\hat{\mu}_{SNV}$	$\hat{\mu}_{DEL}$	$\hat{\mu}_{INS}$	$\hat{\mu}_{TOTAL}$
O1	96	97.25 (0.051)	145.9 (0.76)	1.77 (0.027)	0.25 (0.011)	0.113 (0.0071)	2.13 (0.0296)	1.76 (0.027)	0.25 (0.010)	0.11 (0.007)	2.00 (0.0296)
			600	0.427 (0.0062)	0.061 (0.0027)	0.028 (0.0017)	0.52 (0.0072)	0.43 (0.0062)	0.0605 (0.0027)	0.028 (0.0017)	0.452 (0.0064)
O2	192	97.25 (0.051)	144.75 (1.55)	1.88 (0.032)	0.368 (0.016)	0.23 (0.013)	2.47 (0.046)	1.72 (0.03)	0.34 (0.016)	0.21 (0.018)	2.06 (0.0378)
			620 (0.28)	0.431 (0.0075)	0.084 (0.0034)	0.053 (0.0026)	0.567 (0.0098)	0.40 (0.0074)	0.079 (0.0034)	0.047 (0.0026)	0.443 (0.0083)