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. The classical explanation is a “cost of fidelity”, which posits a tradeoff between selection to reduce deleterious mutations and selection to minimize resources devoted to genome surveillance. An alternative is the “drift barrier”, whereby the mutation rate is under directional selection downward; at equilibrium, the increase in fitness from a further increase in replication fidelity is too small to overcome genetic drift.

We propagated *C. elegans* “mutation accumulation” (MA) lines under minimal selection for ~150 generations, divided each line into two sublines and propagated them for another ~150 generations. The base-substitution rate did not change, but variation among lines increased, as predicted by the CoF. Conversely, the insertion/deletion rate increased, as predicted by the DB. The base-substitution spectrum did not change, whereas the spectra of insertions and deletions did.

Also of interest is the nature of selection on deleterious mutations. If epistasis is synergistic, several evolutionary quandaries could be resolved. To quantify mutational effects on fitness, and to characterize the pattern of epistasis, we assayed competitive fitness 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.

38 The evolution of mutation rate is a longstanding conundrum in evolutionary biology. Because
39 most mutations with effects on fitness are deleterious, why does the mutation rate not evolve to
40 zero[1,2]? The two most compelling hypotheses are (1) the Cost of Fidelity[3] (CoF) and (2) the
41 Drift Barrier[4] (DB). The CoF posits that the optimum mutation rate represents a tradeoff
42 between direct selection to minimize the input of deleterious mutations and indirect selection to
43 minimize the resources devoted to genome surveillance, resulting in stabilizing selection on
44 mutation rate (Figure 1A). The DB posits that directional selection acts to reduce the mutation
45 rate until the fitness increase resulting from a further reduction of the mutation rate is less than
46 the reciprocal of the effective population size (N_e), at which point selection is too weak to
47 counteract genetic drift and the (assumed) mutational bias toward increased mutation rate
48 (Figure 1B).

49 The DB predicts that a decrease in N_e should result in an increase in mutation rate, and
50 vice versa. The CoF makes an analogous prediction: if the mutation rate is at the optimum, a
51 change in N_e, should not result in a directional change in the mutation rate. However,
52 populations at the same optimum should experience a predictable change in the variance where
53 a decrease in N_e should result in increased variance around the optimum, and vice versa.
54 Stabilizing selection comes with a "variance drift barrier" because mutations with fitness effects
55 $s<1/N_e$ will accumulate in the population, irrespective of whether they increase or decrease trait
56 value. However, the prediction of no directional change under relaxed selection is predicated on
57 symmetry of the fitness function around the optimum (Figure 1A). If the fitness function is
58 asymmetric, a population at mutation-selection-drift equilibrium will exhibit a mutational bias,
59 such that more mutations cause the mutation rate to change in the direction of stronger
60 selection than in the direction of weaker selection (Supplementary Figure 1). Without a
61 mutational bias, the population would not be at equilibrium and would be under directional
62 selection.

A second feature of the mutational process of longstanding interest is the distribution of fitness effects (DFE) of new mutations[1,5,6]. 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[7,8].

To investigate these questions in a tractable experimental system, we employed a "second order mutation accumulation" (O2MA) experiment[9-11] using the nematode *C. elegans* (Figure 2A). We initially constructed a set of "first order MA" (O1MA) lines[12-14]. 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.

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, the first-order per-nucleotide SNV rate $\mu_{O1,SNV}$=1.77$\pm$0.038×10^{-9}/generation; the indel rate $\mu_{O1,Indel}$=0.36$\pm$0.019×10^{-9}/generation (Table 1). The total first-order mutation rate (μ_{O1}) is 2.14$\pm$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$\pm$0.034×10^{-9}/generation. The

87 second-order indel rate $\mu_{O2,Indel}$=0.60±0.024×10^{-9}/generation. The total second-order mutation
88 rate μ_{O2}= 2.48±0.046×10^{-9}/generation.
89 False positives (FP) and false negatives (FN) affect the estimated mutation rates in
90 predictable ways (Supplementary Figure 3). FPs artificially inflate the inferred μ_{O2}, and false
91 negatives (FN) artificially deflate μ_{O2} on average. Ignoring FNs results in a conservative test of
92 the hypothesis that $\mu_{O2} > \mu_{O1}$. Because we have only a rough estimate of the number of FPs (≈
93 2 SNVs and 0.33 indels per genome; Supplementary Table 3), we tested the null hypothesis that
94 $\mu_{O2} = \mu_{O1}$ by "titrating" the number of FPs by simulation and asking the question: what is the
95 fewest FPs per O2MA line that would lead us to falsely reject the null hypothesis $\mu_{O2} = \mu_{O1}$ (i.e.,
96 Type I error)? For SNVs, there must be fewer than 0.6 FPs for the observed difference $\Delta\mu_{SNV}$ =
97 0.23×10^{-9}/generation to not be sufficiently explained by FPs (Figure 3A), so we conclude that
98 there is no evidence that $\mu_{O2,SNV} > \mu_{O1,SNV}$. For indels, the same test shows the critical number
99 of FPs per line is 2.7 (Figure 3B), so we conclude that the evidence suggests that $\mu_{O2,Indel} >$
100 $\mu_{O1,Indel}$.
101 Some types of mutations occur independently of genome replication and accrue at an
102 approximately constant rate in chronological time. If the chronological mutation rate remains
103 constant but the generation time increases, the result will be an increase in the per-generation
104 mutation rate. There was little difference in the average number of generations between the O1
105 and O2 lines over the course of the experiment (O1 mean = 145.6 generations; O2 mean =
106 144.2 generations), and the conclusions drawn from mutation rates scaled per-generation apply
107 similarly to mutation rates scaled per-day (Table 1). The number of indels was positively
108 correlated with generations of MA in both O1 ($r_{O1,indel}$=0.29, p<0.004) and O2 lines ($r_{O2,indel}$=0.25,
109 p<0.001). In the O1 lines, the number of SNVs was uncorrelated with the number of
110 generations ($r_{O1,SNV}$=0.10, p>0.32), whereas in the O2 lines the number of SNVs was positively
111 correlated with the number of generations ($r_{O2,SNV}$=0.46, p<<0.0001). The lack of correlation of

112 number of SNVs in the O1MA lines is likely due to lack of power resulting from the smaller
113 sample size and the lack of variation in generations of MA. Taken together, we reject the
114 hypothesis that mutation rate is independent of generation time, but we cannot rule out a
115 contribution of replication-independent mutation.
116
117 *Mutation spectrum*. Consistent with the observed lack of evolution of μ_{SNV}, the SNV mutation
118 spectra did not differ between O1MA and O2MA lines (Monte Carlo Fisher's Exact Test, p >
119 0.12) (Figure 4A,B). G:C→A:T transitions (28.6%) and A:T→T:A transversions (28.1%) were the
120 most common classes of base substitutions, as is typical in *C. elegans* MA experiments[15,16].
121 The indel spectrum, in contrast, was significantly different between O1MA and O2MA
122 lines (Figure 4C,D). In the O1MA lines, deletions outnumbered insertions by nearly 2.2:1,
123 whereas the deletion bias in the O2MA lines was significantly reduced to ~1.6:1 (Fisher's Exact
124 Test (FET), p < 0.0001). Considered separately, the deletion spectrum differed significantly
125 between O1 and O2 lines (FET, p < 0.003), with an increase in the shortest (1 bp) and longest
126 (>10 bp) categories and concomitant decreases in the intermediate categories. The insertion
127 spectrum also differed between O1 and O2 (FET, p < 0.0001), also with increases in the 1 bp
128 and >10 bp categories and decreases in the intermediate categories.
129
130 *Variance in mutation rate*. To begin, we ask: can the observed variance in mutation rate among
131 O1MA lines ($V_{\mu O1}$) be sufficiently explained by sampling variance of a uniform mutational
132 process? For SNVs the answer is yes; $V_{\mu O1,SNV}$ is sufficiently explained by sampling if mutation
133 conforms to a uniform Poisson process (p>0.09, Supplementary Figure 4A). For indels,
134 however, $V_{\mu O1,Indel}$ is too great to be explained by a uniform Poisson process (p<0.02,
135 Supplementary Figure 4B).

136 To extend the analysis to the O2MA lines, FPs must be considered. We assume a
137 liberal average of 3 SNV and 1 indel FPs per sample, leading to a conservative test of the
138 hypothesis of a uniform mutational process. For both SNVs and indels there is more variance
139 among O2MA lines than is consistent with a uniform mutational process (Supplementary Figure
140 5).

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

145 Neither $\mu_{O2,SNV}$ nor $\mu_{O2,Indel}$ is significantly heritable ($h^2_{\mu_{O2,SNV}} = 0.037$, p>0.71; $h^2_{\mu_{O2,Indel}} =$
146 0.089, p>0.38), nor is total mutation rate ($h^2_{\mu_{O2}} = 0.046$, p>0.65). For SNVs, the lack of
147 heritability is consistent with inferred uniformity of the O1MA process. For indels, the simplest
148 explanation is that (*i*) heritability is inherently low, and (*ii*) FPs add additional noise, so 96 O1
149 families provide insufficient power to detect heritability of ~10% over 150 generations. A more
150 speculative possibility is that in the absence of purifying selection, indel mutation rate is an
151 accelerating function of mutation load[17,18].

152

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

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

161 Figure 5, Supplementary Tables 4 and 5). There is a weak negative association between
162 mutation rate and fitness (Supplementary Figure 6).
163 Given that the number of mutations is similar between O1MA and O2MA lines, the
164 implication is that the mean effect on competitive fitness of O1 mutations (u_{O1}) is greater than
165 that of O2 mutations (u_{O2}). To test that hypothesis, we compared a model in which the mean
166 mutational effect u is constrained to remain constant over the entire 300 generations ("uniform
167 effects") to a "two-stage" model in which u_{O1} and u_{O2} are allowed to differ. Models were
168 compared using the Widely Applicable Information Criterion (WAIC) and leave-one-out cross-
169 validation (LOO) to assess model fit and predictive performance[19]. The differences between
170 models are substantial (Δ_{WAIC} = 13.52; Δ_{LOO} = 12.69) and favor the uniform effects model. The
171 point estimate of the mean mutational effect u = -0.0139, with variance $\sigma_g^2 = 1.22\times10^{-4}$ (Figure
172 5).
173
174 **Discussion**
175 Our primary goal was to test empirically the predictions of the two prominent alternative
176 theoretical models of the evolution of mutation rate. The Drift Barrier (DB) model predicts that
177 the mutation rate should increase when selection is relaxed. Explicit in the DB model is the
178 existence of a mutational bias in the direction of increased mutation rate. The Cost of Fidelity
179 (CoF) model posits that mutation rate is under apparent[20] stabilizing selection, whereby direct
180 selection to reduce the input of deleterious mutations is counterbalanced by indirect selection to
181 minimize the resources devoted to genome surveillance. The "strong" form of the CoF is
182 predicated on symmetry of the fitness function around the optimum and predicts that relaxation
183 of selection will not lead to a change in the mean mutation rate, but that the variance in mutation
184 rate will increase. The "weak" form of the CoF, which relaxes the assumption of symmetry of
185 the fitness function, also predicts an increase in the variance in mutation rate upon relaxation of

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

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

199 The findings of this study reinforce and extend previous findings from *C. elegans* and
200 other organisms, in which various types of physiological stress – exogenous and endogenous –
201 increase the indel mutation rate, but have little if any effect on the base substitution mutation
202 rate. Rajaei et al.[15] performed an MA experiment with an N2-derived strain of *C. elegans*
203 carrying a mutation (*Mev-1*) that causes increased cellular oxidative stress. In that experiment,
204 the indel (and specifically, the insertion) rate increased by about 50% over wild-type, whereas
205 the SNV mutation rate decreased slightly. Similarly, MA lines of *Drosophila melanogaster*
206 initiated from strains engineered to carry large-effect deleterious mutations experience a higher
207 rate of short deletions than their wild-type controls, whereas the SNV mutation rates do not
208 differ[21]. *Arabidopsis thaliana* MA lines propagated under benign and stressful conditions of
209 salinity[22] and temperature[23] show large and highly significant increases in indel mutation rates
210 compared to much smaller increases in SNV rates. A similar increase in indel frequency was

211 reported for the green alga *Chlamydomonas reinhardtii* MA lines maintained under salinity
212 stress[24].
213 Our second-order MA experiment has three antecedents. First, Ávila et al.[9] initiated an
214 O2MA experiment from a single high-fitness MA line of *D. melanogaster*. Their study predated
215 economical whole-genome sequencing, but they found that the decline in fitness in the O2MA
216 experiment was greater than in the O1MA experiment. They attributed the increased rate of
217 decline in fitness to an increase in mutation rate, although they could not formally rule out
218 greater mutational effects in the O2MA experiment. The subsequent findings of Sharp and
219 Agrawal[21] add credence to their conclusion. Second, Saxena et al.[10] did an O2MA experiment
220 with *C. elegans*, designed to test the hypothesis that genomes that carry a high mutation load
221 mutate faster than less-loaded genomes[17,18]. The indel rate (specifically the insertion rate)
222 increased significantly in the O2MA lines; contrary to expectation, the O2MA indel rate was
223 greater in the high-fitness (= less-loaded) lines. The SNV rate increased on average, but not
224 significantly. Two of the ten O1MA families (n=5 O2MA sub-lines per O1MA family) evolved
225 SNV O2 mutation rates that were significantly different from the ancestor; one family increased
226 and the other decreased.
227 The third antecedent is the study of Liu and Zhang[11], who brought the power of yeast
228 genetics to bear on the question of mutation rate evolution. They first constructed a set of
229 O1MA lines in which a mismatch repair gene (*MSH-2*) was deleted by CRISPR-Cas9
230 engineering to increase the mutation rate. They then constructed a set of O2MA lines by re-
231 inserting the *MSH-2* gene into the O1MA lines and allowing mutations to accumulate further.
232 They found (*i*) that the O2MA mutation rate increased on average, as predicted by the DB, but
233 also (*ii*) 12/48 O2MA lines evolved a decreased mutation rate. They interpreted the high
234 frequency of lines evolving decreased mutation rate as inconsistent with the DB hypothesis and
235 consistent with the yeast mutation rate being maintained by apparent stabilizing selection (i.e.,
236 the weak CoF). Moreover, they identified a gene with antagonistic pleiotropic effects on cellular

237 physiology (autophagy) and mutation rate, as predicted by the CoF. Their results, if not their
238 conclusions, are in agreement with those of Saxena et al.[10] in *C. elegans*.
239 In 1995, Kondrashov[25] famously posed the question "Why have we not died 100 times
240 over?". Given the sizes of genomes (large) and effective population sizes (much smaller), the
241 continued existence of humans seemed incompatible with the inferred genetic load (~100 lethal
242 equivalents) if the effects of deleterious mutations are independent. The simplest explanation is
243 that the effects of deleterious mutations are not independent, but rather they interact
244 synergistically, i.e., epistasis is negative. However, it is not immediately obvious why
245 deleterious mutations should interact synergistically on average. One possible explanation is
246 that selection is "soft"[26], wherein population absolute fitness is established by resource
247 availability and relative fitness results from competition among individuals for resources[27,28]. If
248 so, it implies that epistasis is more likely to be detected under competitive conditions.
249 This study provides no support for the possibility that deleterious mutations interact
250 synergistically, and in fact the point estimate of the effect of second-order mutations (u_{O2}) is
251 about half that of the first-order effect, suggestive of diminishing returns epistasis. However, we
252 can confidently reject the two-stage model of mutational effects in favor of the one-stage model
253 of constant mutational effects ($\Delta_{\mathrm{WAIC}} \approx 12$). There are obviously many caveats associated with
254 that inference, including environmental context and statistical power. We note that our findings
255 recapitulate those of Peters and Keightley[29] and Katju et al.[30] who looked for epistatic effects of
256 on non-competitive fitness in *C. elegans* and did not find them.
257
258 *Conclusions*. The results of this study are clear: there is an upward mutational bias for indels,
259 whereas there is no such bias for SNV mutations, although the variance in μ_{SNV} increased over
260 the course of the experiment. The discrepancy between evolution of the indel and SNV
261 mutational processes, in this study and others, is intriguing. One possibility is that the genomic
262 mutational target for the control of the indel process is simply larger than that of SNVs, and that

263 if selection was relaxed for long enough that the SNV mutation rate would begin to drift upward.
264 We suspect this is probably true. However, it is worth considering the possibility that the indel
265 and SNV mutational processes are subject to different patterns of selection, and that the DB or
266 weak CoF model applies to indels, whereas the strong CoF model applies to SNVs.

267 What remains to be determined is the nature of the fitness function with respect to
268 mutation rate. Liu and Zhang[11] reported that mutation rate of their yeast O2MA lines is
269 uncorrelated with growth rate. Visual inspection of their plot of growth rate against mutation rate
270 (their Supplementary Figure 2E) reveals a weak positive relationship between mutation rate and
271 growth rate, contrary to the prediction that lines with high mutation rates (mutators) should have
272 lower than average fitness, as expected if mutation rate is under stabilizing selection in the
273 presence of a mutational bias toward increased mutation rate. However, growth rate was
274 measured in only one environment and both mutation rate and fitness in yeast are known to be
275 context dependent[31], so it remains entirely possible that mutation rate in yeast is under global
276 stabilizing selection. What Liu and Zhang clearly did show is (*i*) antimutator mutations are not
277 trivially rare in yeast, but also (*ii*) on average, mutation rate drifts upward when selection is
278 relaxed.

279 The idea that epistasis among deleterious mutations should be synergistic on average
280 has been as influential among evolutionary geneticists as the supporting evidence for it is
281 weak[32]. Beyond the obvious possibility that 50 years of inconclusive evidence is the result of
282 underpowered experiments intended to quantify an inherently noisy trait, the intriguing
283 possibility exists that epistasis itself is not only evolvable[33,34], but that the direction of epistasis
284 favored by selection is dependent on idiosyncratic properties of the population including
285 effective population size and mutation rate[35,36].

286

287 **Methods**

288

289 *Mutation Accumulation (MA) Experiment*

290 We obtained a stock of the EG8072 strain of *C. elegans* (https://cgc.umn.edu/strain/EG8072)
291 from the Caenorhabditis Genetics Center (CGC) at the University of Minnesota in September
292 2018. EG8072 is a derivative of the standard N2 strain, with green (GFP) and red (mCherry)
293 fluorescent reporters stably integrated into the genome. We propagated the founder stock for
294 seven generations by transfer of a single immature hermaphrodite at 4-day intervals, at which
295 point the population is expected to have reached mutation-drift equilibrium ($6N_e$ generations[37]).
296 The population was expanded to large size (thousands of mixed-stage individuals) for 2-3
297 generations and cryopreserved.

298 In November 2018 we thawed an aliquot of the cryopreserved stock, cleaned it by
299 bleaching[38] and initiated 148 replicate populations, each initiated from a single immature
300 hermaphrodite. 48 replicates were expanded to large size (N ≈ a few thousand) for an
301 additional 2-3 generations and cryopreserved in 20 replicate PCR plates. These are the G0
302 ancestral "pseudolines". The remaining 100 populations – the first-order (O1) MA lines - were
303 propagated by transfer of a single immature hermaphrodite at four-day intervals for an additional
304 600 days (Gmax=150 generations). MA lines were maintained under standard conditions: on 35
305 mm NGMA agar plates spotted with 50 µl of an overnight culture of *E. coli* OP50, kept in a
306 Percival incubator set to a constant 20° C. Plates were kept on cafeteria trays, in 6 × 8 arrays,
307 48 plates per tray. Even numbered and odd numbered replicates were kept on different trays.

308 After the leading-generation worm was transferred, the plate it was taken from was
309 moved to a different incubator set to 10° C; this is the (first) backup plate. Worms do not
310 reproduce (or reproduce only very slowly) at 10° C. If at the time of transfer a worm of the
311 leading generation failed to reproduce or was not visibly gravid, it was replaced by a sibling
312 taken from the previous generation plate. If there were offspring on the leading generation plate
313 but no worm had reached at least the L2 stage, the plate was "held over" without transferring a

314 worm. This protocol assures that generation number of a line is known with (near) certainty.
315 Demographic data on backups and holdovers is given in Supplementary Table 6.
316 All lines were cryopreserved at 50 generation intervals. After 150 transfers (600 days,
317 Gmax=150), each O1MA line was replicated into two second-order (O2) MA lines; the two
318 O2MA lines derived from an O1MA line are an "O1MA family". The O2MA lines were
319 propagated as described for an additional ~150 generations (Gmax=306), at which time they
320 were cryopreserved analogously to the G0 pseudolines as described above.
321
322 *Genome Sequencing and Variant Calling*
323 *(i)* DNA Extraction and library preparation. Cryopreserved tubes of worms were thawed and
324 grown for several days on 60 mm NGMA agar plates containing Streptomycin and Nystatin and
325 seeded with 110 µl of overnight culture of the HB101 strain of *E. coli*. When food was nearly
326 exhausted, a small piece (~2 cm^2) of the agar plate was transferred onto a 100 mm NGMA plate
327 and worms were allowed to grow for another 2-3 days. When food was nearly exhausted,
328 worms were collected into 15 ml centrifuge tubes on ice and allowed to settle. Approximately
329 100 µl of settled worms were transferred into a 1.5 ml microcentrifuge tube and stored at -80°C
330 until all samples were ready for DNA extraction, at which time samples were thawed, and
331 genomic DNA was extracted with the DNeasy Kit (Qiagen) following the manufacturer's protocol.
332 Extracted DNA was stored at -20°C prior to library preparation.
333 Frozen DNA samples were thawed and diluted with diH_2O to a concentration of 0.2
334 ng/µl. Tagmentation was performed per manufacturer's instructions (Illumina, Nextera DNA
335 Sample preparation kit, FC-121-1030). Following tagmentation, amplification was performed
336 using custom barcoded IDT primers. Following amplification, a 96-pooled well sample was
337 generated. Once pooled, 170µL of sample was combined with 30µL of 6X loading dye and run
338 on a 2% agarose gel. DNA segments of 300-500 bp were excised from gel. The QIAquick Gel
339 Extraction Kit (QIAquick, 28704) was used to clean and elute DNA from the gel, following

340 manufacturer's instructions. Pooled library samples were quantified using the Qubit HS kit
341 (Qubit, Q33230) and sequenced to ~25X average depth of coverage with 150 bp paired-end
342 Illumina sequencing on a NovaSeq 6000 by Novogene, Inc.
343 *(ii)* Genome Assembly. Raw sequence reads were filtered for quality (MQ≥3) and adapter
344 sequence trimmed with fastp[39]. Reads were then aligned to the *C. elegans* N2 strain reference
345 genome (WS292) using bowtie2[40]. The resulting BAM files were sorted and filtered with
346 Samtools[41] and bamTools[42] for downstream processing. Read group information was added
347 with Picard (http://broadinstitute.github.io/picard). Duplicate reads were identified and removed
348 using MarkDuplicates in GATK[43]. Sequence quality was assessed with FastQC[44] and MultiQC[45]
349 after trimming and alignment and prior to variant calling.
350 *(iii)* Variant Calling Pipeline. Variants were called with the HaplotypeCaller tool in GATK[43], run in
351 GVCF mode. HaplotypeCaller identifies both single nucleotide variants (SNVs) and indels by
352 local de novo assembly of haplotypes in regions with evidence of variation, improving accuracy
353 by reconstructing the sequence context in those regions. GVCFs from all samples were merged
354 with GenomicsDBImport, and joint genotyping was performed using GenotypeGVCFs. We used
355 bcftools[46] to identify homozygous variants unique to a line, which were counted as putative
356 mutations.
357 Variants that met the following three criteria were considered as putative mutations: (1)
358 they were called homozygous; (2) they were present in one and only one O1MA family, and (3)
359 the ancestor genotype is homozygous wild-type. If the variant is present in both O2MA sublines,
360 the mutation is inferred to have occurred in the first 150 generations and be a first-order (O1)
361 mutation. If the variant is present in only one of the two O2MA sublines, it is inferred to have
362 occurred subsequent to the split of the two O2MA sublines and be a second-order (O2)
363 mutation. Comparisons of O1 and O2 mutation rates were restricted to the additional criteria (4)
364 a site must be called in both members of an O1MA family and (5) the site must be called in at
365 least 90% of O1MA families.

366 One O2MA line from O1MA family 9 was not sequenced. The identities of two O2MA
367 lines appear to have been swapped, one from O1MA family 34 and one from family 46. One
368 O1MA line (family 26) evolved an indel mutator phenotype, as evidenced by the >5X greater
369 indel rates of the two O2MA sublines than of the O2MA line with the next highest indel rate
370 (Supplementary Table 1). Except where noted, those families were omitted from further
371 analysis.
372 In the O2MA lines, over 97% of the ~100.2 Mb genome was covered at the liberal 3X
373 coverage criterion (mean= 97.6%, median = 97.8%; Supplementary Table 1). At 3X coverage,
374 the mean and median fractions of the genome included in the O1MA analysis are 97.3% and
375 97.5% respectively. At 10X coverage the O2MA mean and median are 93.5% and 95.8%; the
376 O1MA mean and median are 89.8% and 93.2%.
377
378 *Assessment of falsely inferred variants*
379 (*i*) <u>False positives</u>. EG8072 is derived from the N2 strain, which is the source of the *C. elegans*
380 reference genome[47]. Any difference between our EG8072 ancestor and the reference genome
381 is either (1) a mutation that fixed subsequent to the divergence of EG8072 and the N2 reference
382 strain, (2) a false negative in the N2 reference, or (3) FP in our EG8072 ancestor. In the first
383 two cases, all MA lines derived from the EG8072 ancestor should also have the variant allele.
384 However, if the variant is FP in our EG8072 ancestor, the derived MA lines should all have the
385 reference allele. The EG8072 G0 ancestor was sequenced from three independent Illumina
386 libraries. We estimated the frequency of FPs from each of the three samples of the ancestor.
387 At the liberal (3X) coverage criterion we identified an average of 1662 fixed variants
388 between our G0 ancestor and the reference genome (1676, 1647, and 1664 in the three
389 replicates, respectively; Supplementary Table 3), of which approximately 42% were single
390 nucleotide variants (SNVs) and 58% were insertions or deletions (indels). In the three replicates
391 we identified three (1 deletion, 2 SNVs), three (3 SNVs) and one (1 SNV) FPs respectively. One

392 of the SNVs (a G:C→C:G transversion) was common to all three replicates; the others were
393 only found in one replicate. Of the seven FPs detected at the liberal 3X criterion, six were also
394 found at the stringent 10X criterion. By way of comparison, the same analysis with a different
395 N2 ancestor found one FP (an SNV) out of 1439 variants with 68 MA lines at the 3X criterion[15].
396 (*ii*) Failure to recall (FtR) and false negatives (FN). A mutation can fail to be detected ("failure to
397 recall", FtR) for two reasons. First, a site with a true mutant may not be called (i.e., missing
398 data). Second, a true mutant allele may be mistakenly identified as an ancestral allele; that is a
399 false negative (FN). FNs are a subset of FtR and cause the mutation rate to be underestimated.
400 FtR due to missing data may or may not affect estimation of the mutation rate, depending on
401 whether sites with true mutations are more or less likely to be called than sites without
402 mutations. All else equal, sites with true mutations are less likely to be called than sites without
403 mutations. The reason is that if a sequence with *n* or more mismatches exceeds the cutoff for
404 mapping, a sequence with a mutation has one more mismatch than the same sequence without
405 the mutation. All else also equal, FNs are equally likely to be true O1 mutations or true O2
406 mutations. Since there are twice as many opportunities for O2 as O1 mutations (Supplementary
407 Figure 3), FNs introduce a downward bias in the O2 mutation rate relative to the O1 mutation
408 rate.
409 To assess our ability to recall mutations, we introduced simulated "dummy" mutations at
410 random into the reference genome and analyzed the data using our standard variant-calling
411 pipeline, but with the simulated genome as the reference[15]. The changed site in the reference
412 should appear as a variant in our sequenced sample. For example, suppose the true reference
413 allele at a position is an A and we change it to a C. The site should be called variant (A/A in this
414 example) in our data. If the site is either not covered in the data or is called anything other than
415 a variant homozygote (A/A), we have failed to recall the mutation. If the site is covered in the
416 data but not called a variant homozygote, that is FN.

417 We constructed a pseudo-reference genome by inserting 1200 single-nucleotide variants
418 (SNVs) and 12,000 indels (1–20 bp, size chosen at random) into the *C. elegans* N2 (WS292)
419 reference genome (Supplementary Table 7). First, we randomly inserted "dummy" SNVs and
420 indels into the reference genome in random positions using the R packages GenomicRanges,
421 IRange, and Biostrings and compiled them into a VCF file. The VCF was then sorted and
422 indexed with samtools, and the pseudo-reference genome was produced with GATK
423 FastaAlternateReferenceMaker. The genome assembly and variant calling pipelines were rerun
424 with all the MA lines and the G0 ancestor using this pseudo-reference genome. Line-specific
425 simulated FtR and FN rates and counts are given in Supplementary Table 8.
426 To validate the consistency of the procedure, the original variants (i.e., putative
427 mutations) were re-mapped against the pseudo-reference genome. All sites identified as variant
428 in the original analysis were also called as variant against the pseudo-reference. We repeated
429 the simulation twice independently and obtained consistent results across runs (Supplementary
430 Table 9).
431 At the liberal 3X criterion, the average FtR rate was ~2.2% for SNVs and ~2.5% for
432 indels. The FN rates were 0.21% and 0.53% for SNVs and indels. At the stringent 10X criterion
433 the failure to recall rates were 7.8% and 7.7% for SNVs and indels respectively; the FN rates
434 were 0.084% and 0.12%. All analyses are based on the 3X coverage data; 10X data are
435 reported in the Supplementary Tables.
436
437 *Competitive Fitness Assay*
438 (*i*) Focal worms. 50 G150 O1MA lines and 20 G0 pseudolines (PS lines) were chosen at
439 random for inclusion in assay block 1. Block 2 included one G300 O2MA replicate from each of
440 those same 50 lines and the same 20 PS lines; Block 3 included the other O2MA replicate,
441 along with the same 20 PS lines. At the beginning of each assay (*Day 1*) a cryopreserved
442 aliquot from each line (MA and PS) was thawed, and ~150 µl pipetted onto a 60 mm NGMA

443 plate seeded with the 100 µl of an overnight culture of the OP50 strain of *E. coli* (the thaw plate;
444 Figure 2B) Two days later (*Day 3*), seven immature worms (L3/L4 stage) from each line were
445 picked singly to a 35 mm NGMA plate seeded with 50 µl of OP50. These plates are the parental
446 (P) generation. Each P plate was identified with a random number and arranged in numerical
447 order on cafeteria trays, 48 plates/tray, and placed in a Percival incubator maintained at 20˚C
448 and ambient humidity. Thaw plates were incubated at 10˚C to prevent further reproduction.
449 Four days later (*Day 7*), a single immature (L3/L4) hermaphrodite was picked to a new plate and
450 maintained as before. These plates are the F1 generation. If a P generation worm failed to
451 reproduce, it was replaced in the assay with a sibling from the thaw plate. Four days later (*Day*
452 *11*) the process was repeated for the F2 generation. If an F1 worm failed to reproduce, the
453 replicate was removed from the assay and treated as missing data. Three days later (*Day 14*) a
454 single L1 stage worm was picked from each F2 plate onto the assay plate, along with an L1
455 competitor strain worm.
456 (*ii*) Competitor worms. On *Day 1* of the assay, an aliquot of N2 strain worms was thawed and
457 pipetted onto a seeded 35 mm plate and allowed to reproduce. When the food was exhausted
458 (around *Day 5*), a ~2 cm chunk was transferred to a new 100 mm NGMA plate seeded with 400
459 µl of OP50 and allowed to reproduce. When food was exhausted (around *Day 9*), a chunk was
460 transferred to a new plate. On *Day 13*, worms were washed from the plate and bleached to
461 synchronize the population. Bleached embryos were pipetted in equal aliquots onto four 100
462 mm NGMA plates supplemented with streptomycin and the antifungal agent nystatin and
463 seeded with 400 µl of the HB101 strain of *E. coli*.
464 (*iii*) Assay plates. A single L1 focal worm (MA or G0 pseudoline) and a single L1 competitor
465 worm were picked to a 35 mm NGMA plate supplemented with streptomycin and nystatin and
466 seeded with 50 µl of HB101. Assay plates were identified by random number, placed in random
467 order on trays (48 plates/tray) and incubated at 20°C for eight days (*Day 21*; approximately two
468 overlapping generations), at which point nearly all plates were depleted of bacterial food.

469 Worms were washed from the assay plate in 2 ml of M9 buffer and transferred to a 2 ml deep-
470 well 96-well plate. The plate was centrifuged for 1 min at 1,000 g, and the supernatant was
471 aspirated with an 8-channel strip aspirator set to leave 100 µl in each well. The worms were
472 washed once more by adding 1.4 ml of M9 buffer, centrifuging, and aspirating as before,
473 followed by the addition of 1.4 ml of M9. Worms were mixed by gently pipetting with a 12-
474 channel pipette and a 110 µl sample was transferred into a well in a 96-well, clear-bottom tissue
475 culture plate. 5 mM levamisole was added to each well to immobilize worms prior to imaging.
476 *(iv)* Imaging. Bright-field and GFP fluorescence images (470±20 nm excitation, 525±25 nm
477 emission) were captured at 20x or 30x magnification for each well using an automated
478 epifluorescence microscope (IX-70, Olympus, Pittsburgh, PA, USA) fitted with a CCD camera
479 (Retiga-2000R, QImaging, Surrey, BC, Canada), XYZ stage and focus motors (Prior Scientific,
480 Cambridge, UK), and controlled by Image Pro-plus software (Media Cybernetics, Rockville, MD,
481 USA). 96-well plates were imaged first under bright field (BF) and then under fluorescence
482 (GFP). Images of the same well (BF and GFP) were paired and renamed after checking by eye
483 that the images matched.
484 *(v)* Counting. Worms were counted from stacked images using ImageJ software. We
485 summarize the protocol here; a detailed tutorial of the ImageJ analysis is included as
486 Supplementary Appendix 1.
487 *Step 1*. Paired images were stacked using the “Stacks” option under the “Image” tab.
488 *Step 2*. Images were merged and colorized using the “Merge Channels” tool under the “Image”
489 tab, with fluorescent worms colored differently from non-fluorescent ones.
490 *Step 3.* For best effect, brightness and contrast of GFP images were adjusted by setting
491 “minimum” and “maximum” light low and “contrast” high under the “Image” tab.
492 *Step 4*. If an image was too crowded with worms, the ROI selection tool in the “Analyze” tab was
493 used to crop a sample portion of the stacked images to count. The stacked images were
494 synchronized using Analyze →Tools →Synchronize windows → Synchronize All commands.

495 *Step 5*. Worms were counted by eye using the "Multipoint" tool to mark individual worms, with
496 different marking colors for fluorescent and non-fluorescent worms.
497
498 *Data Analysis*
499 *i) Mutation Rate*. We calculated the per-generation mutation rate, μ_i, for each MA line *i* as $\mu_i =$
500 $m_i/n_i t_i$, where *m* is the number of mutations, *n* is the number of callable sites, and *t* is the number
501 of generations of MA[48]. First-order and second-order mutation rates were calculated separately.

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

507 To proceed, we employed a parametric bootstrap approach. We assume that mutation
508 can be modeled as a Poisson process. We further assume that there is a uniform mutation rate
509 $\mu = \mu_{O1} = \mu_{O2}$, i.e., the mutation rate does not change over the course of the experiment.
510 Because FPs necessarily inflate μ_{O2} relative to μ_{O1} and FNs do the opposite on average,
511 ignoring FNs leads to a conservative test of the prior hypothesis that $\mu_{O2} > \mu_{O1}$. To begin, we
512 set μ= to the point estimate of μ_{O1}. At each generation, an MA line accumulates mutations in its
513 genome drawn from a Poisson distribution with parameter $\lambda_m = \mu L$, where μ is the per-
514 generation mutation rate and *L* is the size of the *C. elegans* genome. The total number of
515 mutations in line *i*, $m_i = \lambda_m t_i$, where t_i is the number of generations of MA for line *i*. At
516 Gmax=150, the O1MA line is split into the two O2MA sublines, and the mutational process is
517 continued for the remaining Gmax=150 generations. Each mutation was accepted ("called") or
518 rejected with binomial probability equal to the proportion of the genome covered in the line
519 (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[10].

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[49] (Equation 6.4 in Ref. 49). 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[50]. 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:

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

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[51] 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.

596 **Data Availability**

597 Raw sequence data are deposited in the NCBI Sequence Read Archive BioProject
598 #PRJNA1089216. Competitive fitness data are deposited in Dryad UPON ACCEPTANCE.
599 Code and unprocessed image files are available at https://github.com/Baer-Group/O2MAno2.

600

601 **Acknowledgments**

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

606

607 **Figure Legends**

608

609 **Figure 1**. Theoretical models of mutation rate evolution. (**A**) Cost of Fidelity. (**B**) Drift Barrier.
610 **Figure 2**. (**A**) Schematic depiction of the MA experiment. Short black lines depict ancestral G0
611 "pseudolines" (PS), blue lines depict O1MA lines, orange lines depict O2MA lines. (**B**)
612 Schematic depiction of the competitive fitness assay. Label "LX" refers to "Line X" (MA or G0),
613 label LX.1,…LX.7 refers to replicates 1-7 of a line in the assay. See Methods for details.
614 **Figure 3**. Distribution of simulated values of $\Delta\mu=\mu_{O2}-\mu_{O1}$ with increasing number of false
615 positives. Red dashed line is the observed value. "FP cutoff" is the minimum number of false
616 positives necessary for the data to be consistent with the hypothesis $\Delta\mu=0$. (**A**) SNVs (**B**)
617 Indels.
618 **Figure 4**. Type-specific mutation rates and spectra. (**A**) SNV mutation rates (**B**) SNV frequency
619 spectrum (**C**) Indel mutation rates (**D**) Indel frequency spectrum.
620 **Figure 5**. Distribution of posterior probabilities of mutational effects u on competitive fitness. (**A**)
621 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

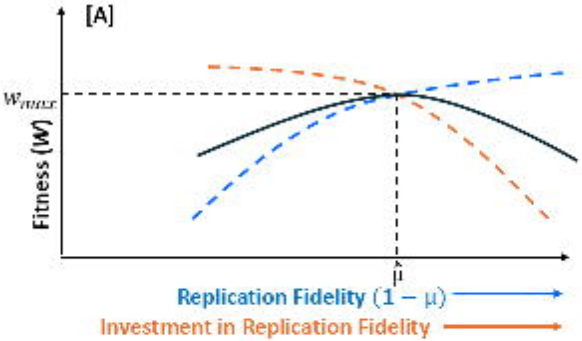
1. Fisher, R.A. *The Genetical Theory of Natural Selection*, (Clarendon Press, Oxford, 1930).
2. Sturtevant, A.H. Essays on evolution. 1. On the effects of selection on mutation rate. *Quarterly Review of Biology* **12**, 467–477 (1937).
3. Kimura, M. On evolutionary adjustment of spontaneous mutation rates. *Genetical Research* **9**, 23–27 (1967).
4. Lynch, M. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* **180**, 933–943 (2008).
5. Charlesworth, B. The good fairy godmother of evolutionary genetics. *Curr Biol* **6**, 220 (1996).
6. Eyre-Walker, A. & Keightley, P.D. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**, 610–618 (2007).
7. Barton, N.H. A general model for the evolution of recombination. *Genetical Research* **65**, 123–144 (1995).
8. Otto, S.P. The evolutionary enigma of sex. *American Naturalist* **174**, S1–S14 (2009).
9. Ávila, V. *et al.* Increase of the spontaneous mutation rate in a long-term experiment with *Drosophila melanogaster*. *Genetics* **173**, 267–277 (2006).
10. Saxena, A.S., Salomon, M.P., Matsuba, C., Yeh, S.D. & Baer, C.F. Evolution of the mutational process under relaxed selection in *Caenorhabditis elegans*. *Molecular Biology and Evolution* **36**, 239–251 (2019).
11. Liu, H.X. & Zhang, J.Z. The rate and molecular spectrum of mutation are selectively maintained in yeast. *Nature Communications* **12**(2021).
12. Keightley, P.D. & Caballero, A. 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 (1997).
13. Vassilieva, L.L. & Lynch, M. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151**, 119–129 (1999).
14. Baer, C.F. *et al.* 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 (2005).
15. Rajaei, M. *et al.* 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 (2021).
16. Konrad, A., Brady, M.J., Bergthorsson, U. & Katju, V. Mutational landscape of spontaneous base substitutions and small indels in experimental *Caenorhabditis elegans* populations of differing size. *Genetics* **212**, 837–854 (2019).
17. Agrawal, A.F. Genetic loads under fitness-dependent mutation rates. *Journal of Evolutionary Biology* **15**, 1004–1010 (2002).
18. Shaw, F.H. & Baer, C.F. Evolutionary consequences of fitness-dependent mutation rate in finite populations. *Journal of Evolutionary Biology* **24**, 1677–1684 (2011).

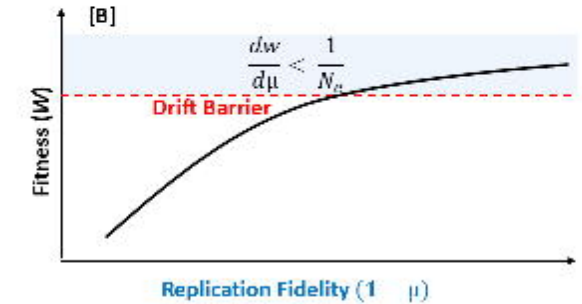
19. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**, 1413–1432 (2017).
20. Kondrashov, A.S. & Turelli, M. Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. *Genetics* **132**, 603–618 (1992).
21. Sharp, N.P. & Agrawal, A.F. Low genetic quality alters key dimensions of the mutational spectrum. *PLoS Biology* **14**(2016).
22. Jiang, C.F. *et al*. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Research* **24**, 1821–1829 (2014).
23. Belfield, E.J. *et al*. Thermal stress accelerates *Arabidopsis thaliana* mutation rate. *Genome Research* (2020).
24. Hasan, A. *et al*. Salt stress alters the spectrum of de novo mutation available to selection during experimental adaptation of Chlamydomonas reinhardtii. *Evolution* **76**, 2450–2463 (2022).
25. Kondrashov, A.S. Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *Journal of Theoretical Biology* **175**, 583–594 (1995).
26. Wallace, B. Polymorphism, population size, and genetic load. in *Population Biology and Evolution* (ed. Lewontin, R.C.) 87–108 (Syracuse University Press, Syracuse, NY, 1968).
27. Peck, J. & Waxman, D. Mutation and sex in a competitive world. *Nature* **406**, 399–404 (2000).
28. Agrawal, A.F. & Whitlock, M.C. Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? *Trends in Ecology & Evolution* **25**, 450–458 (2010).
29. Peters, A. & Keightley, P. A test for epistasis among induced mutations in Caenorhabditis elegans. *Genetics* **156**, 1635–1647 (2000).
30. Katju, V., Packard, L.B., Bu, L., Keightley, P.D. & Bergthorsson, U. Fitness decline in spontaneous mutation accumulation lines of Caenorhabditis elegans with varying effective population sizes. *Evolution* **69**, 104–116 (2015).
31. Liu, H.X. & Zhang, J.Z. Yeast spontaneous mutation rate and spectrum vary with environment. *Current Biology* **29**, 1584–1591 (2019).
32. de Visser, J., Cooper, T. & Elena, S. The causes of epistasis. *Proceedings of the Royal Society B-Biological Sciences* **278**, 3617–3624 (2011).
33. Azevedo, R.B.R., Lohaus, R., Srinivasan, S., Dang, K.K. & Burch, C.L. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* **440**, 87–90 (2006).
34. Desai, M., Weissman, D. & Feldman, M. Evolution can favor antagonistic epistasis. *Genetics* **177**, 1001–1010 (2007).
35. Sydykova, D., LaBar, T., Adami, C. & Wilke, C. Moderate Amounts of Epistasis are Not Evolutionarily Stable in Small Populations. *Journal of Molecular Evolution* **88**, 435–444 (2020).

36. Gros, P., Le Nagard, H. & Tenaillon, O. The Evolution of Epistasis and Its Links With Genetic Robustness, Complexity and Drift in a Phenotypic Model of Adaptation. *Genetics* **182**, 277–293 (2009).
37. Lynch, M. & Hill, W.G. Phenotypic evolution by neutral mutation. *Evolution* **40**, 915–935 (1986).
38. Stiernagle, T. Maintenance of *C. elegans*. in *WormBook* (ed. Community, T.*C.e.*R.) (2006).
39. Chen, S.F., Zhou, Y.Q., Chen, Y.R. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, 884–890 (2018).
40. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–U54 (2012).
41. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P. & Marth, G.T. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).
43. McKenna, A. *et al*. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010).
44. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
45. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
46. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–93 (2011).
47. Ichikawa, K. *et al*. CGC1, a new reference genome for *Caenorhabditis elegans*. *Genome Research* **35**, 1902–1918 (2025).
48. Denver, D.R. *et al*. 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 (2009).
49. Gillespie, J.H. *Population Genetics: A Concise Guide*, (Johns Hopkins University Press, Baltimore, MD, 2004).
50. Stearns, F., Zhou, J. & Fenster, C. Scaling the fitness effects of mutations with respect to differentially adapted Arabidopsis thaliana accessions under natural conditions. *Evolution* **79**, 951–961 (2025).
51. Salvatier, J., Wiecki, T.V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* **2**, e55 (2016).

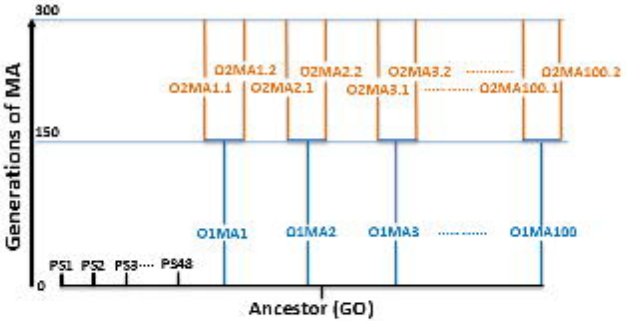
Table 1. Summary statistics (SEM in parentheses). N_{Lines} is the number of MA lines, $\bar{t}$ is the average number of generations of MA (top row) and number of days of MA (bottom row). Values of $\bar{\mu}$ are mutation rates ($\times 10^9$) calculated from the observed number of mutations without 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)

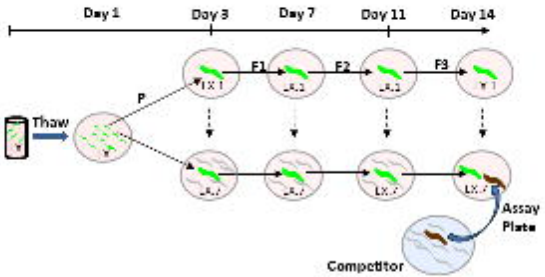

[A]
w_{max}
Fitness (W)
μ
Replication Fidelity (1 − μ)
Investment in Replication Fidelity


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

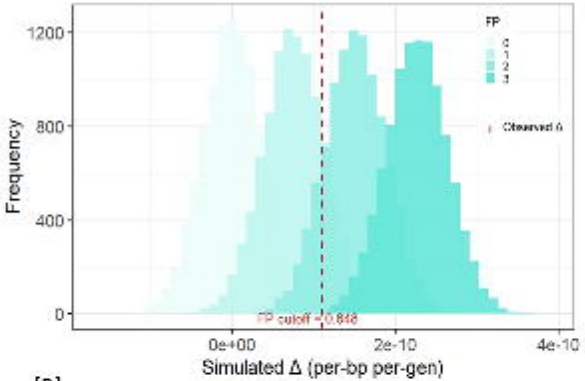
[A]


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)

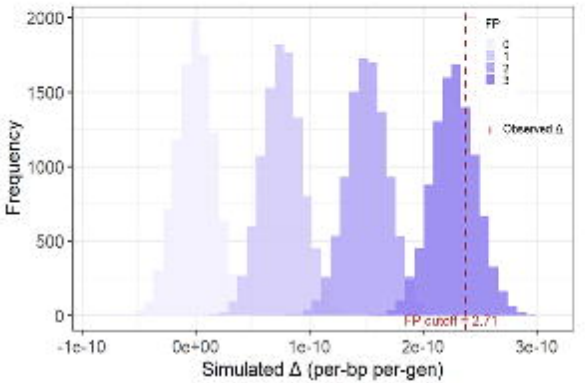
[B]

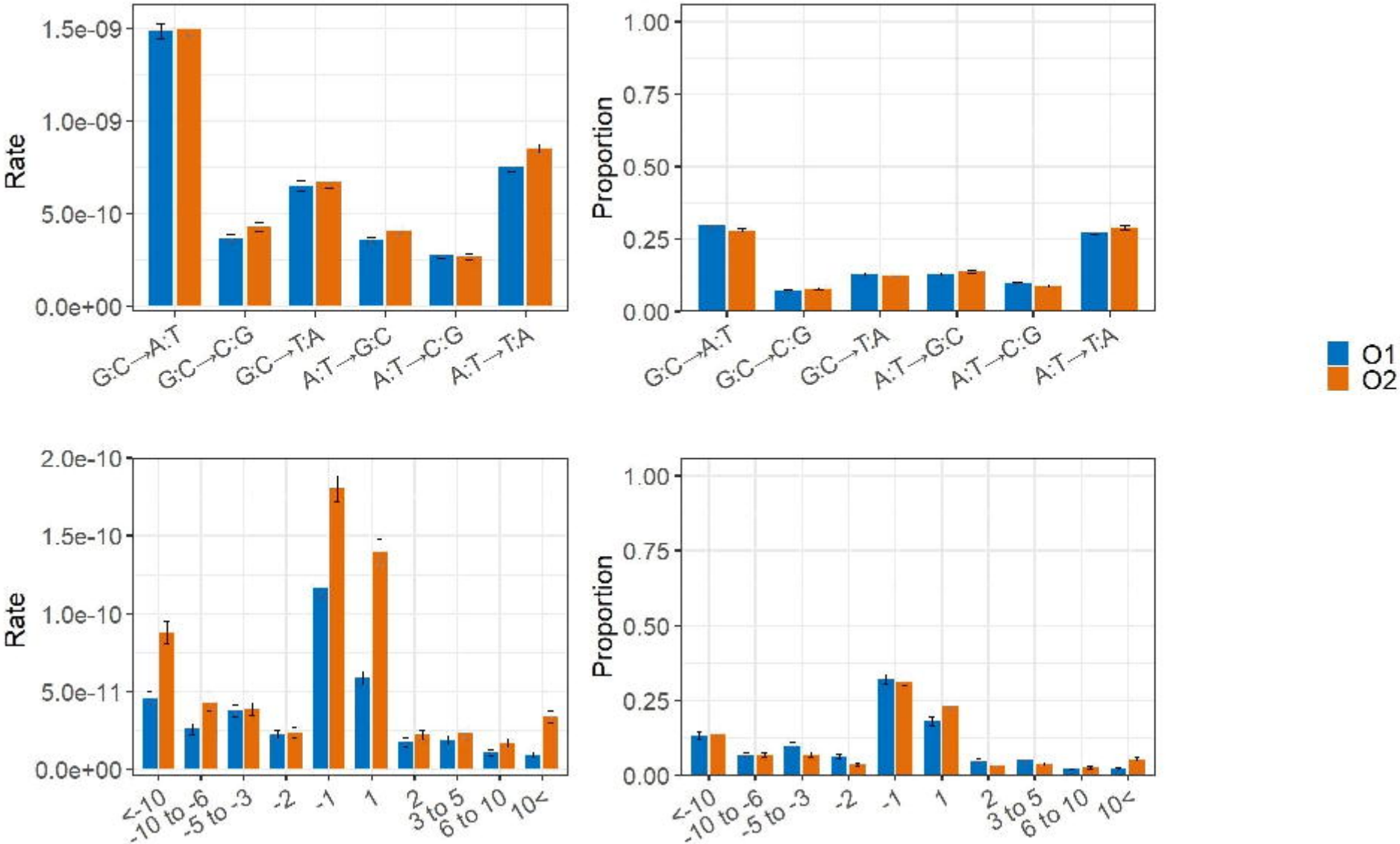

Day 1
Day 3
Day 7
Day 11
Day 14
F1
F2
F3
P
Thaw
Assay Plate
Competitor

[A]

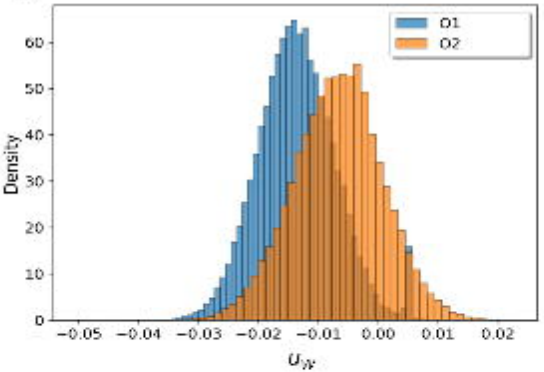

FP
0
1
2
3
Observed Δ
Frequency
1200
800
400
0
FP cutoff = 0.848
0e+00
2e-10
4e-10
Simulated Δ (per-bp per-gen)

[B]

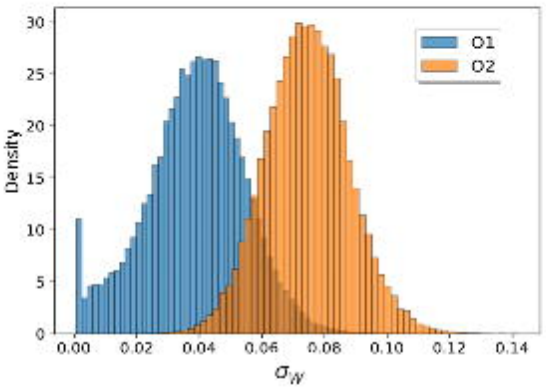

FP
0
1
2
3
Observed Δ
Frequency
2000
1500
1000
500
0
FP cutoff = 2.71
-1e-10
0e+00
1e-10
2e-10
3e-10
Simulated Δ (per-bp per-gen)


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
$\sigma_{/N}$