Evolution of the rate, molecular 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 32611 USA
2- Department of Biology, Johns Hopkins University, Baltimore, MD 21218 USA
3- Department of Ecology and Genetics, Uppsala University, ,Norbyvägen 18D, 752 36 Uppsala, Sweden
4- Current Address: School of Public Health, Yale University, New Haven, CT 06510 USA
5- Current Address: Regeneron, Inc., Tarrytown, NY 10591 USA
6- University of Florida Genetics Institute, Gainesville, FL 32611 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

The rate, molecular spectrum, and fitness effects of mutations vary at all levels of the biological hierarchy, from within individual genomes to among taxonomic domains. Understanding the evolutionary factors underpinning that variation is of fundamental importance to biology. Accurate quantification of the properties of mutations requires that other evolutionary forces, especially natural selection, be minimized as much as possible.

To investigate the evolution of the mutational process in *C. elegans*, we propagated a set of 100 “first order” mutation accumulation (O1MA) lines under minimal selection for ~150 generations, divided each O1MA line into two “second order” MA (O2MA) lines and propagated them for another ~150 generations, at which time the genome of each O2MA line was sequenced, and a subset of 50 O1MA families was assayed for competitive fitness.

Over the course of the experiment, the mean nucleotide substitution mutation rate did not change, but the variance increased. In contrast, the indel mutation rate increased significantly. The two types of mutations fulfill the predictions of different theoretical models for the evolution of mutation rate. These results reinforce previous findings that the rate of indels is more sensitive to endogenous stress than the rate of nucleotide substitutions.

Several evolutionary quandaries could be resolved if deleterious mutations interact synergistically (negative epistasis). Evidence for synergistic epistasis is famously inconclusive, although there is reason to think it may be more detectable under competitive conditions. However, a model of constant mutational effects on competitive fitness explains the results significantly better than a model including epistasis.

Introduction

Evolution begins with mutation and would soon end if it stopped. Understanding the mutational process is thus of fundamental importance to all areas of evolutionary biology. Two distinct, but intertwined, lines of inquiry are (1): what factors influence the rate and molecular spectrum of mutation? and (2) what factors influence the effects of mutations on the phenotype and on fitness? In any case, to make progress it is necessary to consider a sufficiently large set of mutations unbiased by the effects of other evolutionary forces, natural selection in particular. In principle, this can be done either by investigating the properties of extremely rare segregating variants (Messer 2009, Schraiber et al. 2025), which most likely arose recently as new mutations and have been minimally scrutinized by natural selection, or by means of a laboratory mutation accumulation (MA) experiment, in which populations are manipulated in such a way as to minimize the efficacy of natural selection (Mukai 1964, Katju and Bergthorsson 2019). Each method has strengths and drawbacks. It is possible to characterize the sequence features of vastly more segregating variants than could ever be accumulated in an MA experiment, and those mutations integrate over genotype and environment in (recent) time and space. However, the phenotypic effects of very rare segregating variants are usually not measurable, and the fitness effects are estimatable only indirectly and are confounded with the effects of demography (Johri et al. 2022).

MA experiments, on the other hand, have the significant limitation that only a comparatively small number of mutations can be accumulated on at most a few genetic backgrounds in at most a few environmental contexts that are assuredly unnatural for the organism. However, MA experiments come with two major advantages: first, that the experimenter controls demography (and thus the efficacy of selection) and the environment, and second, that the effects on phenotype and fitness can be measured directly, albeit in the lab environment and (usually) only in aggregate.

It is by now well-established that the rate and spectrum of mutation vary even within species and populations (Sasani et al. 2022, Bergeron et al. 2023, Lynch et al. 2023, Wang and Obbard 2023), and the causes of that variation are beginning to be disentangled (Liu and Zhang 2019, Wei et al. 2022, Sasani et al. 2024). Most studies designed to parse the causes of variation in the mutational process have been done in microbes, primarily for reasons of tractability, although relevant studies have been done in a variety of multicellular organisms, including humans (Carlson et al. 2018, Carlson et al. 2020, Garcia-Salinas et al. 2025). A few broad-scale patterns have emerged, among them (*i*) the per-genome, per-generation mutation rate (U) varies inversely with population size and body size, at least at large phylogenetic scale (Lynch et al. 2023), (*ii*) the per-nucleotide, per-generation mutation rate (μ) varies inversely with genome size in microbes ("Drake's Rule", Drake 1991) but positively with genome size in multicellular eukaryotes (Lynch et al. 2016), and (*iii*) the mutation rate often increases under conditions of physiological stress (Ram and Hadany 2012, Williams and Foster 2012, Liu and Zhang 2019). Obviously, however, much remains to be learned about the causes of variability in the mutational process.

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). One feature of the DFE is certain, and another nearly so. First, it is beyond doubt that most mutations with non-trivial effects on fitness are deleterious. Second, in diploid organisms, there is abundant evidence for a negative relationship between the strength of selection and the degree of dominance: on average, the more deleterious the mutation, the closer it is to being completely recessive (Simmons and Crow 1977, Di and Lohmueller 2024). Beyond that, the picture gets murky.

For several reasons, it is of great interest to know how mutations, especially deleterious mutations, interact *on average*. Whether the average 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), as well as the long-term existence of species with modest effective population sizes (e.g., humans, Kondrashov 1995, Charlesworth 2013b). The evident long-term advantage of sex and recombination becomes easier to explain if deleterious mutations interact synergistically, as does our continued existence.

For related reasons, it is also of considerable interest to understand the relationship between environmental stress and the fitness effects of mutations. There is a longstanding general intuition that mutations become more deleterious under stress, but as trenchantly pointed out by Agrawal and Whitlock (2010), there is no a priori theoretical justification for that intuition, although they convincingly argue that deleterious effects are likely to be magnified under competitive conditions.

To experimentally investigate the evolution of the mutational process in the (near) absence of natural selection, we employ what we refer to as a "second-order" MA experiment (Figure 1A). The O2MA design lets us quantify the evolution of the rate and spectrum of mutation, as well as the evolution of mutational effects on fitness. A difference in the mutational properties (rate, spectrum, and/or fitness effects) between mutations accumulated during the first-order and second-order phases of the experiment are indicative of evolution of some element of the mutational process. The resulting data allow us to address several fundamental questions in evolutionary biology. Of specific interest are (*i*) the cumulative directional effect of new (anti)mutator alleles (i.e., the mutational bias for mutation rate), and (*ii*) the rate of input of new genetic variance for mutation rate (mutators and antimutators). The relative frequency of (anti)mutators has important implications for the evolution of mutation rate (Drake 1993, Lynch 2011), and it is commonly assumed that mutator alleles arise by spontaneous mutation more often than antimutators (e.g., Lynch et al. 2016). However, there is little quantitative information with respect to the rate and bias of spontaneous mutations affecting mutation rate, especially in

multicellular organisms, and a recent study in yeast reported a surprisingly high frequency of spontaneous antimutators (Liu and Zhang 2021).

Also of interest is (*iii*) the relationship between mutation rate and fitness. If the optimum mutation rate is zero (Sturtevant 1937, Kimura 1967, Lynch 2008), the expectation is that fitness will always be negatively correlated with mutation rate. However, if the mutation rate is under stabilizing selection with an optimum greater than zero (Kimura 1967, Fitzsimmons et al. 2018), the implication is that genotypes with intermediate mutation rates will have higher fitness than genotypes with more extreme mutation rates. Further, if the mutation rate is condition-dependent, the mutation rate may itself be an increasing function of the mutation load, which has implications for both the evolution of sex and recombination (Agrawal 2002) and the survival of small populations (Shaw and Baer 2011).

Finally, we are interested in (*iv*) the average interaction between the accumulated mutations with respect to fitness. The O2MA design allows us to quantify the average mutational effect as a function of the mutation load, and thus the average direction of epistasis. The average epistatic effect of new mutations has been of interest to population geneticists for over fifty years (Mukai 1969), but the empirical evidence is frustratingly inconclusive (de Visser et al. 2011). However, there is some reason to believe that epistasis is magnified under competitive conditions (Peck and Waxman 2000), and most empirical investigations of mutational epistasis have not measured competitive fitness.

Items (*i*) and (*ii*) are addressed by means of whole-genome sequencing; items (*iii*) and (*iv*) are addressed using experimental assays of competitive fitness. We elaborate on each of these topics in the Discussion.

Methods and Materials

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 1.

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) DNA Extraction and library preparation. 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. Adapter sequences were trimmed from the raw sequence reads 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. Sites were filtered using GATK VariantFiltration to retain genotypes with a minimum coverage depth of either DP ≥ 3 or DP ≥ 10. These two filtered datasets were then analyzed in parallel in all subsequent steps. 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 (G0) 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 O2MA lines in an O1MA family, (5) the site must be called in at least 90% of O1MA families, and (6) variants within 100 bp of another variant ("multi-nucleotide variants", MNV) in the same O1MA family were omitted from the analysis of SNVs and indels. MNVs are analyzed as a separate category of mutation.

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 extreme 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 2). Investigation of the mutations in family 26 revealed two potential O1 mutator mutations in the genes *chd-3* and *mut-16*. *chd-3* is a chromodomain helicase binding protein involved in the repair of meiotic double-strand breaks (Turcotte et al. 2018); *mut-16* is involved in (among other things) nuclear siRNA suppression of TEs (Zhang et al. 2011). 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%). 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% (Supplementary Table 2).

Assessment of falsely inferred variants

(*i*) False positives. 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 (FN) in the N2 reference, or (3) a false positive (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 derived from the same extracted DNA sample. 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 40% were single nucleotide variants (SNVs), 37% were insertions, and 22% were deletions. Approximately 20% of those variants are multi-nucleotide variants (MVNs), defined as any variant occurring within 100 bases of another variant. Proportions of variants in an MNV that are SNVs, insertions, and deletions are 53%, 28% and 20%, respectively. In the three G0 replicates we identified three (1 deletion, 2 SNVs), three (3 SNVs) and one (1 SNV) FPs respectively. One of the FP SNVs (a G:C→C:G transversion) was common to all three replicates; the others were only found in one replicate. The sole deletion is in an MNV within 10 bp of an SNV in the same replicate line. 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 in one replicate of the G0 ancestor and 68 MA lines at the 3X criterion (Rajaei et al. 2021).

(*ii*) Failure to recall (FtR) and false negatives (FN). 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 "true" 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 bias 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 in any sequenced O2MA line an FN is equally likely to be an O1 or O2 mutation (Figure 2) FNs do not introduce a systematic bias (assuming equal O1 and O2 mutation rates).

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 an 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 4). 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 5.

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 6).

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 O1MA families and the same 20 PS lines; Block 3 included the other O2MA replicate from each of the same 50 O1MA families, 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 1B) 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 Material.

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 $\mu_i = m_i/n_i t_i$, 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 (Figure 2).

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}, the analysis must account for FPs. 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 2). 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 $\mu_{O2} = \mu_{O1}$.

The preceding analysis is a hypothesis test. To put confidence intervals on the estimates of μ_{O1} and μ_{O2}, we used a non-parametric bootstrap, as follows. For μ_{O1}, the procedure is:

Step 1. Resample the 96 O1MA lines with replacement, holding the numbers of O1 mutations (m_{i1}), callable sites (n_i), and O1 generations (t_{i1}) equal to the observed for each line *i*.

Step 2. To each line, add an additional *FtR* mutations where *FtR* ~Poisson(λ_{FtR}) where $\lambda_{FtR} = 0.022 \times m_{i1}$ for SNVs and $0.025 \times m_{i1}$ for indels. These are the mutations which we failed to recall.

Step 3. Calculate the estimated mutation rate μ_{O1} as before, where $m_1^* = m_1 + FtR$.

Step 4. Repeat the above procedure 10000 times; the upper and lower 2.5% of bootstrap replicates establish the 95% confidence interval on μ_{O1}.

For μ_{O2}, the procedure is:

Step 1. Resample the 192 O2MA lines with replacement, holding the numbers of O2 mutations m_{i2}, callable sites (n_i), and O2 generations (t_{i2}) equal to the observed for each line *i*.

Step 2. For each line *i*, subtract FP_i false positives from m_{i2}, where FP_i ~Poisson(λ_{FP}) with λ_{FP} = 2 for SNVs and 0.33 for indels. Since the observed number of false positive insertions is zero, but the true probability of a false positive insertion must be non-zero, indels were assigned as deletions or insertions with probability equal to the observed (deletion) bias. .

Step 3. To each line *i*, add an additional *FtR* mutations where *FtR* ~Poisson(λ_{FtR}) where $\lambda_{FtR} = 0.022 \times m_{i2}$ for SNVs and $0.025 \times m_{i2}$ for indels. These additional mutations are O2 failutre to recall.

Step 4. Calculate the estimated mutation rate μ_{O2} as before, where $m_2^* = m_2 - FP + FN$.

Step 5. Repeat the above procedure 10000 times; the upper and lower 2.5% of bootstrap replicates establish the 95% confidence interval on μ_{O2}.

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, we can – with a caveat - estimate the mutational variance (V_M) for mutation rate. V_M is defined as the per-generation increase in genetic variance for a trait resulting from new mutations and is a fundamental parameter in quantitative genetics (Clayton and Robertson 1955, Lynch and Hill 1986). The among-O1MA family component of variance, V_L, represents the cumulative genetic variance introduced over 150 generations of accumulated mutations . In principle, we would want to estimate V_M for mutation rate from each O2MA line prior to the O2MA phase, because mutations accumulating in the O2 phase will necessarily contribute to the variation among O1MA families after the O2 phase (i.e., at generation 300). That can be easily done in microbes, in which mutation rate can be measured over one round of replication by means of a fluctuation test (Luria and Delbrück 1943), but would require 50% more sequencing in our experiment. Instead, we partition the variance in mutation rate at generation 300 into within- and among-O1MA family components. The among-O1MA family component of variance is V_L; the residual within-family (i.e., between O2MA line) component of variance is V_E. $V_M = V_L/2t$, where t is the number of generations of MA (≈ 150) (Lynch and Walsh 1998, p. 330). The mutational heritability, $h_m^2 = \frac{V_M}{V_E}$. The caveat is that V_M estimated in this way is a lower bound, because of the contribution of O2 mutations to the total variance. Variance components were calculated using restricted maximum likelihood (REML) in a general linear model (GLM) as implemented in the MIXED Procedure of SAS v.9.4, with μ_{O2} as the dependent variable, O1MA family as a random effect and O2MA line nested within O1MA family as a repeated measure. Statistical significance of V_M was assessed by likelihood ratio test (LRT) of the full model against a model with V_L constrained to zero.

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).

We assume that the fitness effects of O1 and O2 mutations are drawn from arbitrary DFEs with mean μ_1 and μ_2, and standard deviations σ_1^2 and σ_2^2, respectively. Given that the numbers of observed O1 and O2 mutations are fairly large, assuming mutations combine additively to determine fitness, we can ignore the higher moments of the DFEs and fairly approximate $g_{1,i}$ and $g_{2,i}$ using normal distributions by the Central Limit Theorem (see

Supplemental Figure 1 in Stearns et al. (2025) for a comparison between the distributions of the sum of mutational effects sampled from arbitrary DFEs and their Gaussian approximations). Specifically, we sampled the vector of genetic effects (g_i) from a multivariate Gaussian distribution

$$\boldsymbol{g} \sim N(\mu_1 \boldsymbol{C}_1 + \mu_2 \boldsymbol{C}_2, \sigma_1^2 \boldsymbol{G}_1 + \sigma_2^2 \boldsymbol{G}_2),$$

where $\boldsymbol{g}$ is a vector containing the genetic effects for all observations. $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are diagonal matrices

$$C_{1,ii} = n_{1,i} \text{ and } C_{2,ii} = n_{2,i},$$

where $n_{1,i}$ and $n_{2,i}$ are the number of O1 and O2 mutations in individual i. $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ account for the covariance in genetic effects between observations, with diagonal terms

$$G_{1,ii} = n_{1,i} \text{ and } G_{2,ii} = n_{2,i}.$$

The off-diagonal terms record the number of shared mutations, such that

$$G_{1,ij} = n_{1,i},$$

if i and j share a common ancestor in generation 150, and $G_{1,ij} = 0$ otherwise. Similarly,

$$G_{2,ij} = n_{2,i},$$

if they share a common ancestor in generation 300.

We fit two competing models under this formulation. In the one-stage ("uniform effects") model, we assume the O1 and O2 mutations have a common DFE, that is, $\mu_1 = \mu_2$, and $\sigma_1^2 = \sigma_2^2$. In the two-stage model, we assume the DFEs of the O1 and O2 mutations are different and independently infer the parameters μ_1 and μ_2, σ_1^2 and σ_2^2.

In addition to one-stage and two-stage model, we also fit a model where we independently estimate the effects of SNVs and indels, while assuming constant DFEs for O1 and O2 mutations. The sampling procedure was a straightforward modification of the steps listed above.

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 λ = 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.

Results

Mutation rate. Over the first ~150 generations we observed 2897 putative mutations (2400 SNVs, 341 deletions, and 156 insertions; Supplementary Tables 2 and 7, Supplementary Figure 1). Ignoring false inferences for the moment, the observed first-order per-nucleotide SNV rate $\mu_{O1,SNV}$=1.77×10^{-9}/generation; the indel rate $\mu_{O1,Indel}$=0.36×10^{-9}/generation. The total first-order mutation rate μ_{O1}=2.13×10^{-9}/generation. In the second ~150 generations of the experiment we observed 6631 total variants (5031 SNVs, 986 deletions, 614 insertions). The observed second-order SNV rate $\mu_{O2,SNV}$=1.88×10^{-9}/generation; the second-order indel rate

$\mu_{O2,Indel}$=0.60×10^{-9}/generation. The total second-order mutation rate μ_{O2}= 2.47×10^{-9}/generation.

False positives (FP) and false negatives (FN) affect the estimated mutation rates in predictable ways (Figure 2). FPs artificially inflate the inferred μ_{O2}, whereas FNs do not introduce bias, on average. 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 2.71 (Figure 3B). The probability of observing no more than one FP in three replicates (see Methods and Supplementary Table 3) if the true Poisson mean is 2.71 is approximately 0.0028, so we conclude that the evidence suggests that $\mu_{O2,Indel} > \mu_{O1,Indel}$.

Estimates of mutation rates corrected for false inferences and 95% confidence intervals are given in Table 1.

Some types of mutations occur independently of genome replication and accrue at an approximately constant rate in chronological time (Gao et al. 2016, Gao et al. 2019). 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 = 145.9 generations; O2 mean = 144.8 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 4). 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 4). 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 (>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.

Multi-nucleotide variants (MNVs). It is known that mutations occur in clusters more often than expected by chance (Schrider et al. 2011), and that short tandem repeats tend to accumulate SNVs ("imperfect microsatellites") in addition to indels (Saxena et al. 2019). We (arbitrarily) define an MNV cluster as all variants within 100 bp of their nearest neighbor, although >90% of all MNV mutations occur within 20 bp of their nearest neighbor. To quantify the rate and

spectrum of MNV mutations and to account for the possibility that MNVs may result in mapping errors that incorrectly identify one O1 MNV as two different O2 mutations, one in each O2 MA sibling line, we re-analyzed the data with and without MNVs (Supplementary Table 8). At the liberal 3X coverage criterion, more O2 SNVs occur in MNVs than O1 SNVs (7.2% O2 vs. 4.4% O1; Pearson's chi-square = 20.3, P<0.0001). In contrast, the proportion of insertions in MNVs is greater in O1 than O2 (34.0% O1, 22.5% O2; Pearson's chi-square = 8.82, P<0.01). The proportion of deletions in MNVs does not differ significantly between O1 and O2 (16.2% O2 vs 12.6% O1; Pearson's chi-square = 2.56, P>0.10). The results are similar at the more stringent 10X coverage criterion (Supplementary Table 8). Summary statistics of mutation rate with MNVs considered as a separate class of mutation are given in Supplementary Table 9.

To account for the possibility that O1 mutations in MNVs may be mis-mapped as two different O2 mutations, we re-ran the FP "titration" analysis described above with mutations in MNVs omitted. The conclusions do not change: the SNV rate does not change significantly from O1 to O2, whereas the indel rate increases by about 60% from O1 to O2 (critical FP = 2.24, Poisson P=0.009; Supplementary Figure 2). The greater proportion of SNVs in MNVs from O1 to O2 can be explained by the increased rate of indel mutations (especially insertions) from O1 to O2 if the mutational process that generates indels also tends to generate associated SNVs.

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 3A). For indels, however, $V_{\mu O1,Indel}$ is too great to be explained by a uniform Poisson process (p<0.02, Supplementary Figure 3B).

To extend the analysis to the O2MA lines, FPs must be accounted for. We assume an average of 2 SNV and 0.33 indel FPs per sample. For both SNVs and indels there is more variance among O2MA lines than is consistent with a uniform mutational process (Supplementary Figure 3).

Estimates of mutational variances (V_M) for mutation rate do not differ significantly from zero for SNVs, indels, or total mutation rate (Supplementary Table 10). REML estimates of the among-O1MA family component of variance (V_L) for SNVs and total mutation rate are 0; the point-estimate of the mutational heritability (h_m^2) for indel rate is on the order of 10^{-4}/generation, roughly an order of magnitude less than for typical quantitative traits (Houle et al. 1996, Conradsen et al. 2022). The lack of significant V_M for SNV mutation rate is consistent with the inferred uniformity of the O1MA process. For indels, the simplest explanation for the discrepancy between the two ways of quantifying variance ($V_{\mu O1}$ vs. V_M) is that the effects of mutations accumulated in the O2 phase largely swamp those of the O1 mutations, removing the signal of heritability (although when the strong mutator O1MA family 426 is included in the analysis, h_m^2 for indel rate ≈ 0.1/generation, which we take as a cautionary note).

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 (see Methods for explanation).

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 4, Supplementary Tables 11 and 12). There is a weak negative association between mutation rate and fitness (Supplementary Figure 5).

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 σ_g^2 = 1.22×10^{-4} (Figure 5).

Given the different mutational properties of SNVs and indels, we attempted to assess potential differences in fitness effects between the two types of mutations by fitting a model allowing separate DFE parameters for the two mutation classes. However, the WAIC values for the one-DFE and two-DFE models are virtually identical, providing no evidence for different fitness effects between SNVs and indels. Although it is possible that the average fitness effects of SNVs and indels do not differ, the result is more simply explained by the limited statistical power resulting from the smaller number of mutations in each class.

The O2MA design permits us to address the question: is the O2 mutation rate a function of O1 fitness? The absence of significant mutational heritability for mutation rate does not preclude *a priori* a relationship between O1 fitness and O2 mutation rate, unless O1 mutation rate strongly predicts O1 fitness, which it does not (Supplementary figure 5). However, O2 mutation rate (SNV, indel, and total mutation rate) is not correlated with O1 fitness (Pearson's r<0.09, p>0.59 in all cases.

Discussion

Our foremost interest is in the evolution of the mutation rate. Mutator alleles (e.g., a defective allele of a DNA repair gene) increase the mutation rate; antimutator alleles (e.g., an allele at a regulatory sequence that increases the expression level of a DNA repair gene) decrease the mutation rate. Intuition suggests that spontaneous mutator mutations should be more common than antimutator mutations, on the grounds that biological systems are well-adapted and a random change is more likely to make a well-adapted system worse than it is to improve it (Fisher 1930, Drake 1993, Lynch 2012). That intuition leads to a clear prediction: if spontaneous mutations are allowed to accumulate unimpeded by natural selection, the mutation rate should increase over time. However, one can imagine alternative scenarios in which the converse is true. For example, if (*i*) mutations accrue approximately constantly in chronological time and (*ii*) growth rate (and thus generation time) is a function of metabolic rate and (*iii*) deleterious mutations decrease metabolic rate on average, then a random mutation might be expected to decrease the per-generation mutation rate. Or along the same lines, if metabolic byproducts (e.g., free radicals) are an important cause of mutation (Martin and Palumbi 1993, Gillooly et al. 2005) and if deleterious mutations decrease metabolic rate on average, the mutation rate might decrease as deleterious mutations accumulate.

Given the importance of mutation to all areas of biology, and the vast trove of theoretical investigations of the evolution of mutation rate (Drake et al. 1998, Lynch et al. 2016), there have been surprisingly few empirical studies that have attempted to quantify the effects of spontaneous mutations themselves on the rate and spectrum of mutation. In the first such study of which we are aware, Ávila et al. (2006) initiated an O2MA experiment from a single high-fitness MA line of *Drosophila 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 (i.e., synergistic epistasis).

Sharp and Agrawal (2016) compared the spontaneous mutation rates between a wild-type strain of *D. melanogaster* and derivative strains engineered to carry mutations of large effect, to test the hypothesis that genomes that carry a high load of deleterious mutations mutate faster than less-loaded genomes (Agrawal 2002, Shaw and Baer 2011).. They found that mutation rate of the mutant strains was higher, due to an increased frequency of short deletions, which they attributed to differences in the repair of double-strand breaks.

Saxena et al. (2019) did an O2MA experiment with *C. elegans*, also designed to investigate the relationship between initial mutation load and mutation rate. The indel rate (specifically the insertion rate) increased significantly in the O2MA lines; contrary to expectation, however, 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 lines per O1MA family) evolved SNV O2 mutation rates that were significantly different from the ancestor; one family increased and the other decreased.

Liu and Zhang (2021) 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 that the O2MA mutation rate increased on average, but a surprisingly high proportion of the O2MA lines (12/48) evolved a decreased mutation rate. Moreover, they identified a gene with antagonistic pleiotropic effects on cellular physiology (autophagy) and mutation rate.

Classical theory predicts that, in the presence of recombination, direct selection favors a reduction in the mutation rate except under special circumstances (an example of the “reduction principle”, Liberman and Feldman 1986), which leads directly to “Sturtevant’s Paradox”: why does the mutation rate not evolve to zero? (Sturtevant 1937). There are two competing (but not mutually exclusive) hypotheses for why the mutation rate does not evolve to zero. In historical

chronological order, the Cost of Fidelity (CoF) hypothesis (Supplementary Figure 6A) posits that the mutation rate is under stabilizing selection, with the optimum established by a tradeoff between direct selection to reduce the input of deleterious mutations and indirect selection to minimize the resources (time, energy, molecules) devoted to replication fidelity and genome surveillance (Kimura 1967). At an equilibrium mutation rate established by the CoF, the per-generation fitness cost of an antimutator must equal the cumulative per-generation fitness cost of the deleterious mutations spawned by an equivalent mutator. In contrast, the Drift Barrier hypothesis (Supplementary Figure 6B) posits that the mutation rate is always under directional selection downwards to reduce the input of deleterious mutations (per the reduction principle), but that at some point 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 (Lynch 2008, 2011).

The existence of a mutational bias means that the trait is under directional selection in the direction opposed to the bias in the vicinity of the ancestral (i.e., unmutated) mean (Charlesworth 2013a, Lynch and Menor 2025). However, there is no information about the fitness function outside the range of the data. A mutational bias, as we observed for indels, fulfills a strong prediction of the DB hypothesis, but does not refute the existence of a non-zero fitness optimum, as predicted by the CoF. In contrast, an increase in the trait variance in the absence of a mutational bias, as we observed for SNVs, is not consistent with the DB hypothesis, and fulfills a key prediction of the CoF (i.e., some deleterious mutations will increase the rate and decrease the cost, and other deleterious mutations will decrease the rate and increase the cost).

Our findings about the evolution of the mutation spectrum 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. (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. As noted above, MA lines of *D. 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). Several studies have investigated the mutational properties of *C. elegans* in MA experiments done with elements of the DNA repair system knocked out, which has the overall effect of greatly increasing the indel rate, but the bias is typically toward deletions rather than insertions (Meier et al. 2014, Meier et al. 2018, Volkova et al. 2020).

Our second main question of interest concerns the evolution of mutational effects on fitness. Thirty years ago, Kondrashov (1995) famously posed the question “Why have we not died 100 times over?”. Given the size of the human 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). 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 ($\Delta_{WAIC} \approx 12$). 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 new mutations on non-competitive fitness in *C. elegans* and did not find them.

Conclusions and Future Directions. The results of this study are clear: there is an upward mutational bias for indels over the course of 300 generations, 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, with the indel rate under directional selection downward (the DB) and the SNV rate under stabilizing selection (the CoF). We are unaware of any study in which any element of the mutation rate has drifted significantly downward.

What remains to be determined is the nature of the fitness function on the downward (antimutator) side of the ancestral mean mutation rate. The obvious way to address that question is to assess fitness as a function of mutation rate over the course of one generation, from a source population with significant variation in mutation rate. That can be done straightforwardly in microbes by means of a fluctuation test. Liu and Zhang (2021) did exactly

that and reported that mutation rate of their yeast O2MA lines is uncorrelated with growth rate. Unfortunately, multicellular organisms do not lend themselves to fluctuation tests, and a suitably powered one-generation experiment using our methods would be extremely labor intensive and require a great deal of sequencing. However, genome sequencing is getting ever cheaper, and advances in high-throughput phenotyping are promising (Mok et al. 2023, O'Brien et al. 2025, Tintori et al. 2025).

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). One obvious possibility is that 50+ years of inconclusive evidence is the result of underpowered experiments intended to quantify an inherently noisy trait. It is now possible to disentangle the confounding between mutation rate and mutational effects that bedeviled early studies (e.g., Mukai 1969, Ávila et al. 2006), and here too, high-throughput phenotyping holds promise. However, 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). The truth may already be out there.

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. We thank the reviewers and the Associate Editor for their many insightful comments and suggestions. Support was provided by National Institutes of Health awards GM127433 (CFB, VK, and ECA) and GM154908 (JZ).

References Cited

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

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

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

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

Azevedo, R. B. R., R. Lohaus, S. Srinivasan, K. K. Dang, and C. L. Burch. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature **440**:87–90.

Barnett, D. W., E. K. Garrison, A. R. Quinlan, M. P. Strömberg, and G. T. Marth. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics **27**:1691–1692.

Barton, N. H. 1995. A general model for the evolution of recombination. Genetical Research **65**:123–144.

Belfield, E. J., C. Brown, Z. J. Ding, L. Chapman, M. Luo, E. Hinde, S. W. van Es, S. Johnson, Y. Ning, S. J. Zheng, A. Mithani, and N. P. Harberd. 2020. Thermal stress accelerates *Arabidopsis thaliana* mutation rate. Genome Research.

Bergeron, L., S. Besenbacher, J. Zheng, P. Li, M. Bertelsen, B. Quintard, J. Hoffman, Z. Li, J. Leger, C. Shao, J. Stiller, M. Gilbert, M. Schierup, and G. Zhang. 2023. Evolution of the germline mutation rate across vertebrates. Nature **615**:285–291.

Carlson, J., W. DeWitt, and K. Harris. 2020. Inferring evolutionary dynamics of mutation rates through the lens of mutation spectrum variation. Current Opinion in Genetics & Development **62**:50–57.

Carlson, J., A. E. Locke, M. Flickinger, M. Zawistowski, S. Levy, R. M. Myers, M. Boehnke, H. M. Kang, L. J. Scott, J. Z. Li, S. Zöllner, D. Absher, H. Akil, G. Breen, M. Burmeister, S. Cohen-Woods, W. G. Iacono, J. A. Knowles, L. Legrand, Q. Lu, M. McGue, M. G. McInnis, C. N. Pato, M. T. Pato, M. Rivera, J. L. Sobell, J. B. Vincent, S. J. Watson, and B. C. The. 2018. Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. Nature Communications **9**:3753.

Charlesworth, B. 1996. The good fairy godmother of evolutionary genetics. Current biology : CB **6**:220.
Charlesworth, B. 2013a. Stabilizing Selection, Purifying Selection, and Mutational Bias in Finite Populations. Genetics **194**:955–971.
Charlesworth, B. 2013b. Why we are not dead 100 times over. Evolution **67**:3354–3361.
Chen, S. F., Y. Q. Zhou, Y. R. Chen, and J. Gu. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**:884–890.
Clayton, G., and A. Robertson. 1955. Mutation and quantitative variation. American Naturalist **89**:151–158.
Conradsen, C., M. W. Blows, and K. McGuigan. 2022. Causes of variability in estimates of mutational variance from mutation accumulation experiments. Genetics **221**.
de Visser, J., T. Cooper, and S. Elena. 2011. The causes of epistasis. Proceedings of the Royal Society B-Biological Sciences **278**:3617–3624.
Denver, D. R., P. C. Dolan, L. J. Wilhelm, W. Sung, J. I. Lucas-Lledo, D. K. Howe, S. C. Lewis, K. Okamoto, W. K. Thomas, M. Lynch, and C. F. Baer. 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., D. Weissman, and M. Feldman. 2007. Evolution can favor antagonistic epistasis. Genetics **177**:1001–1010.
Di, C., and K. Lohmueller. 2024. Revisiting Dominance in Population Genetics. Genome Biology and Evolution **16**.
Drake, J. W. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proceedings of the National Academy of Sciences of the United States of America **88**:7160–7164.
Drake, J. W. 1993. General antimutators are improbable. Journal of Molecular Biology **229**:8–13.
Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics **148**:1667–1686.
Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics **32**:3047–3048.
Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics **8**:610–618.
Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.
Fitzsimmons, W. J., R. J. Woods, J. T. McCrone, A. Woodman, J. J. Arnold, M. Yennawar, R. Evans, C. E. Cameron, and A. S. Lauring. 2018. A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLoS Biology **16**.
Gao, Z. Y., P. Moorjani, T. A. Sasani, B. S. Pedersen, A. R. Quinlan, L. B. Jorde, G. Amster, and M. Przeworski. 2019. Overlooked roles of DNA damage and maternal age in generating human germline mutations. Proceedings of the National Academy of Sciences of the United States of America **116**:9491–9500.
Gao, Z. Y., M. J. Wyman, G. Sella, and M. Przeworski. 2016. Interpreting the Dependence of Mutation Rates on Age and Time. PLoS Biology **14**.

Garcia-Salinas, O., S. Hwang, Q. Huang, R. Sanghvi, D. Malawsky, J. Kaplanis, M. Neville, F. Day, R. Rahbari, A. Scally, and H. Martin. 2025. The impact of ancestral, genetic, and environmental influences on germline de novo mutation rates and spectra. Nature Communications **16**.

Gillooly, J. F., A. P. Allen, G. B. West, and J. H. Brown. 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. Proceedings of the National Academy of Sciences of the United States of America **102**:140–145.

Gros, P., H. Le Nagard, and O. Tenaillon. 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., J. Lachapelle, S. El-Shawa, R. Potjewyd, S. Ford, and R. Ness. 2022. Salt stress alters the spectrum of de novo mutation available to selection during experimental adaptation of *Chlamydomonas reinhardtii*. Evolution **76**:2450–2463.

Houle, D., B. Morikawa, and M. Lynch. 1996. Comparing mutational variabilities. Genetics **143**:1467–1483.

Ichikawa, K., M. Shoura, K. Artiles, D. Jeong, C. Owa, H. Kobayashi, Y. Suzuki, M. Kanamori, Y. Toyoshima, Y. Iino, A. Rougvie, L. Wahba, A. Fire, E. Schwarz, and S. Morishita. 2025. CGC1, a new reference genome for *Caenorhabditis elegans*. Genome Research **35**:1902–1918.

Jiang, C. F., A. Mithani, E. J. Belfield, R. Mott, L. D. Hurst, and N. P. Harberd. 2014. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. Genome Research **24**:1821–1829.

Johri, P., A. Eyre-Walker, R. Gutenkunst, K. Lohmueller, and J. Jensen. 2022. On the prospect of achieving accurate joint estimation of selection with population history. Genome Biology and Evolution **14**.

Katju, V., and U. Bergthorsson. 2019. Old Trade, New Tricks: Insights into the Spontaneous Mutation Process from the Partnering of Classical Mutation Accumulation Experiments with High-Throughput Genomic Approaches. Genome Biology and Evolution **11**:136–165.

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

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

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

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

Liberman, U., and M. W. Feldman. 1986. Modifiers of mutation rate: A general reduction principle. Theoretical Population Biology **30**:125–142.

Liu, H. X., and J. Z. Zhang. 2019. Yeast spontaneous mutation rate and spectrum vary with environment. Current Biology **29**:1584–1591.

Liu, H. X., and J. Z. Zhang. 2021. The rate and molecular spectrum of mutation are selectively maintained in yeast. Nature Communications **12**.

Luria, S. E., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28**:491–511.

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. 2012. Evolutionary layering and the limits to cellular perfection. Proceedings of the National Academy of Sciences **109**:18851–18856.

Lynch, M., M. S. Ackerman, J. F. Gout, H. Long, W. Sung, W. K. Thomas, and P. L. Foster. 2016. Genetic drift, selection and the evolution of the mutation rate. Nature Reviews Genetics **17**:704–714.

Lynch, M., F. Ali, T. Lin, Y. Wang, J. Ni, and H. Long. 2023. The divergence of mutation rates and spectra across the Tree of Life. EMBO reports **24**.

Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. Evolution **40**:915–935.

Lynch, M., and S. Menor. 2025. The divergence of mean phenotypes under persistent Gaussian selection. Genetics **229**.

Lynch, M., and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, MA.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences of the United States of America **90**:4087–4091.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20**:1297–1303.

Meier, B., S. L. Cooke, J. Weiss, A. P. Bailly, L. B. Alexandrov, J. Marshall, K. Raine, M. Maddison, E. Anderson, M. R. Stratton, A. Gartner, and P. J. Campbell. 2014. *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. Genome Research **24**:1624–1636.

Meier, B., N. V. Volkova, Y. Hong, P. Schofield, P. J. Campbell, M. Gerstung, and A. Gartner. 2018. Mutational signatures of DNA mismatch repair deficiency in *C. elegans* and human cancers. Genome Research **28**:666–675.

Messer, P. W. 2009. Measuring the Rates of Spontaneous Mutation From Deep and Large-Scale Polymorphism Data. Genetics **182**:1219–1232.

Mok, C., M. Xiao, Y. Wan, W. Zhao, S. Ahmed, R. Luallen, and A. Reinke. 2023. High-throughput phenotyping of infection by diverse microsporidia species reveals a wild *C. elegans* strain with opposing resistance and susceptibility traits. Plos Pathogens **19**.

Mukai, T. 1964. Genetic structure of natural populations of *Drosophila melanogaster*. 1. Spontaneous mutation rate of polygenes controlling viability. Genetics **50**:1–19.

Mukai, T. 1969. Genetic structure of natural populations of *Drosophila melanogaster*. 7. Synergistic interaction of spontaneous mutant polygenes controlling viability. Genetics **61**:749–761.

O'Brien, T. J., I. L. Barlow, L. Feriani, and A. E. X. Brown. 2025. High-throughput tracking enables systematic phenotyping and drug repurposing in *C. elegans* disease models. Elife **12**:RP92491.

Otto, S. P. 2009. The evolutionary enigma of sex. American Naturalist **174**:S1–S14.

Peck, J., and D. Waxman. 2000. Mutation and sex in a competitive world. Nature **406**:399–404.

Peters, A., and P. Keightley. 2000. A test for epistasis among induced mutations in *Caenorhabditis elegans*. Genetics **156**:1635–1647.

Rajaei, M., A. S. Saxena, L. M. Johnson, M. C. Snyder, T. A. Crombie, R. E. Tanny, E. C. Andersen, J. Joyner-Matos, and C. F. Baer. 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.

Ram, Y., and L. Hadany. 2012. The evolution of stress-induced hypermutation in asexual populations. Evolution **66**:2315–2328.

Salvatier, J., T. V. Wiecki, and C. Fonnesbeck. 2016. Probabilistic programming in Python using PyMC3. PeerJ Computer Science **2**:e55.

Sasani, T., D. Ashbrook, A. Beichman, L. Lu, A. Palmer, R. Williams, J. Pritchard, and K. Harris. 2022. A natural mutator allele shapes mutation spectrum variation in mice. Nature **605**:497–502.

Sasani, T., A. Quinlan, and K. Harris. 2024. Epistasis between mutator alleles contributes to germline mutation spectrum variability in laboratory mice. Elife **12**.

Saxena, A. S., M. P. Salomon, C. Matsuba, S. D. Yeh, and C. F. Baer. 2019. Evolution of the mutational process under relaxed selection in *Caenorhabditis elegans*. Molecular Biology and Evolution **36**:239–251.

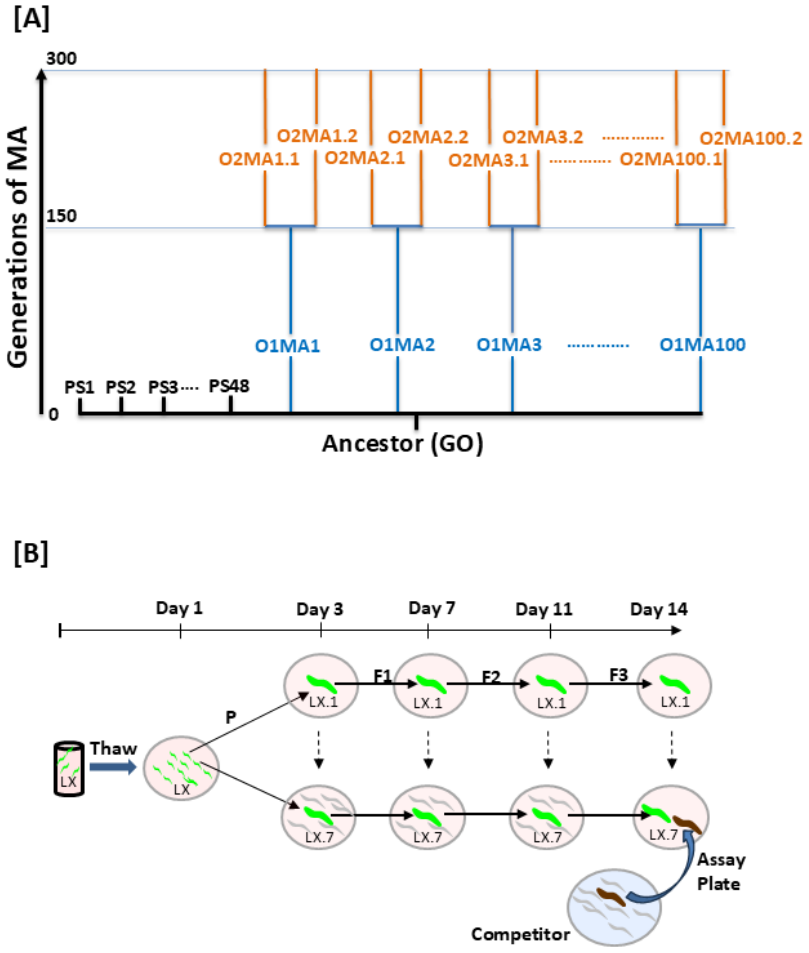
Schraiber, J., J. Spence, and M. Edge. 2025. Estimation of demography and mutation rates from one million haploid genomes. American Journal of Human Genetics **112**:2152–2166.

Schrider, D., J. Hourmozdi, and M. Hahn. 2011. Pervasive Multinucleotide Mutational Events in Eukaryotes. Current Biology **21**:1051–1054.

Sharp, N. P., and A. F. Agrawal. 2016. Low genetic quality alters key dimensions of the mutational spectrum. PLoS Biology **14**.
Shaw, F. H., and C. F. Baer. 2011. Evolutionary consequences of fitness-dependent mutation rate in finite populations. Journal of Evolutionary Biology **24**:1677–1684.
Simmons, M. J., and J. F. Crow. 1977. Mutations affecting fitness in Drosophila populations. Annual Review of Genetics **11**:49–78.
Stearns, F., J. Zhou, and C. Fenster. 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* T. *C. e.* R. Community, editor. Wormbook.
Sturtevant, A. H. 1937. Essays on evolution. 1. On the effects of selection on mutation rate. Quarterly Review of Biology **12**:467–477.
Sydykova, D., T. LaBar, C. Adami, and C. Wilke. 2020. Moderate Amounts of Epistasis are Not Evolutionarily Stable in Small Populations. Journal of Molecular Evolution **88**:435–444.
Tintori, S., D. Caglar, and M. Rockman. 2025. A multigenerational population-growth assay to capture subtle fitness phenotypes in *C. elegans* and other nematodes. Genetics **230**.
Turcotte, C. A., S. A. Sloat, J. A. Rigothi, E. Rosenkranse, A. L. Northrup, N. P. Andrews, and P. M. Checchi. 2018. Maintenance of Genome Integrity by Mi2 Homologs CHD-3 and LET-418 in *Caenorhabditis elegans*. Genetics **208**:991–1007.
Vehtari, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing **27**:1413–1432.
Volkova, N. V., B. Meier, V. Gonzalez-Huici, S. Bertolini, S. Gonzalez, H. Vohringer, F. Abascal, I. Martincorena, P. J. Campbell, A. Gartner, and M. Gerstun. 2020. Mutational signatures are jointly shaped by DNA damage and repair. Nature Communications **11**.
Wallace, B. 1968. Polymorphism, population size, and genetic load. Pages 87–108 *in* R. C. Lewontin, editor. Population Biology and Evolution. Syracuse University Press, Syracuse, NY.
Wang, Y., and D. J. Obbard. 2023. Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis. Evolution Letters **7**:216–226.
Wei, W., W. Ho, M. Behringer, S. Miller, G. Bcharah, and M. Lynch. 2022. Rapid evolution of mutation rate and spectrum in response to environmental and population-genetic challenges. Nature Communications **13**.
Williams, A. B., and P. L. Foster. 2012. Stress-Induced Mutagenesis. EcoSal Plus **5**:10.1128/ecosalplus.1127.1122.1123.
Zhang, C., T. Montgomery, H. Gabel, S. Fischer, C. Phillips, N. Fahlgren, C. Sullivan, J. Carrington, and G. Ruvkun. 2011. mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America **108**:1201–1208.

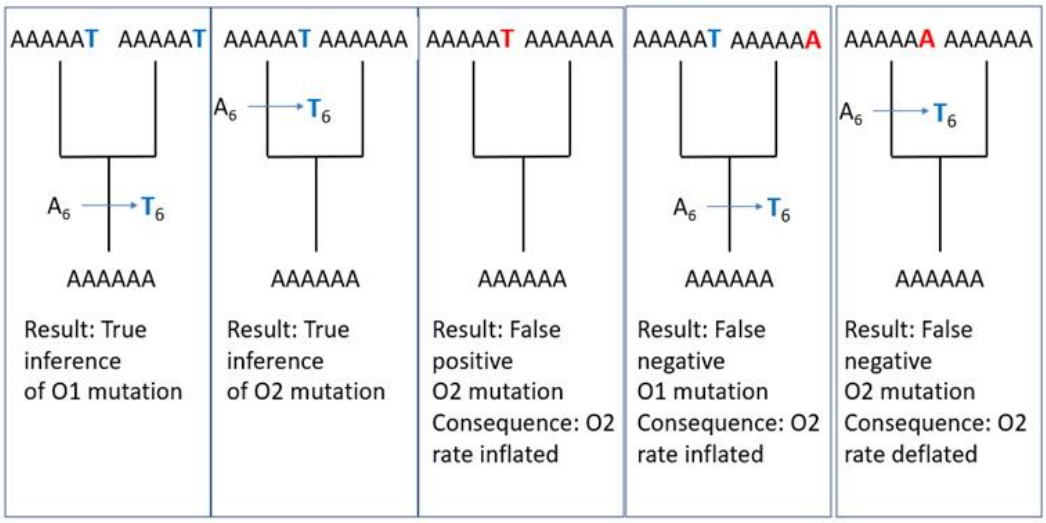
Figure Legends

Figure 1 caption. (**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.



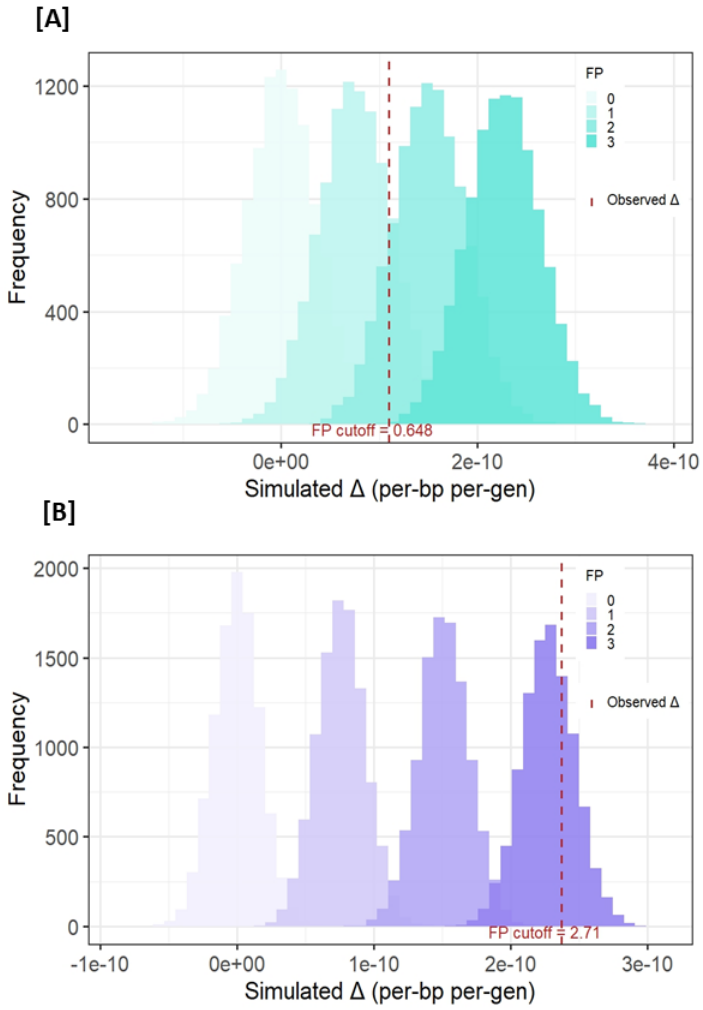
Alt Text: Graphical depiction of the MA experiment and the fitness assay. Panel (A) shows the schematic diagram of the MA experiment; panel (B) shows a schematic diagram of the fitness assay.

Figure 2. 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 at sequence position 6.



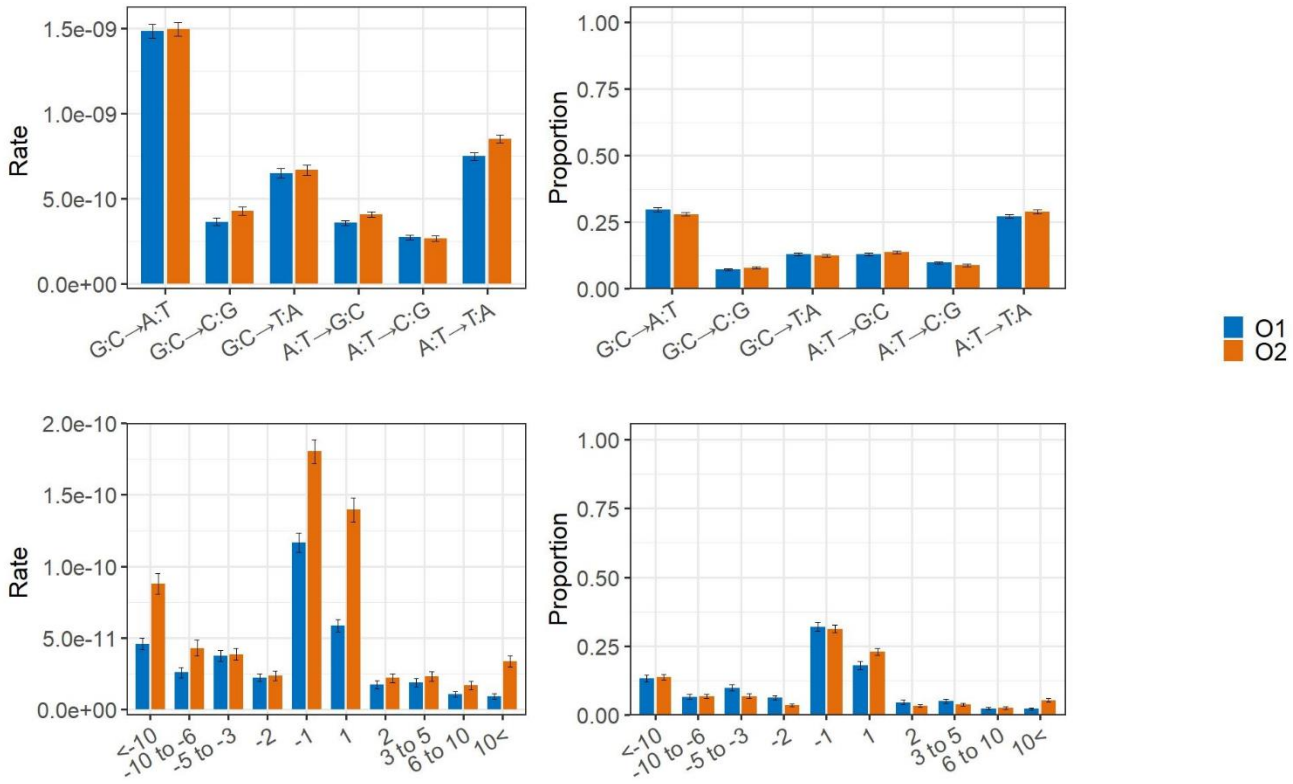
Alt Text: Graphical depiction of the effects of falsely inferred sequences on the inferred mutation rate.

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.



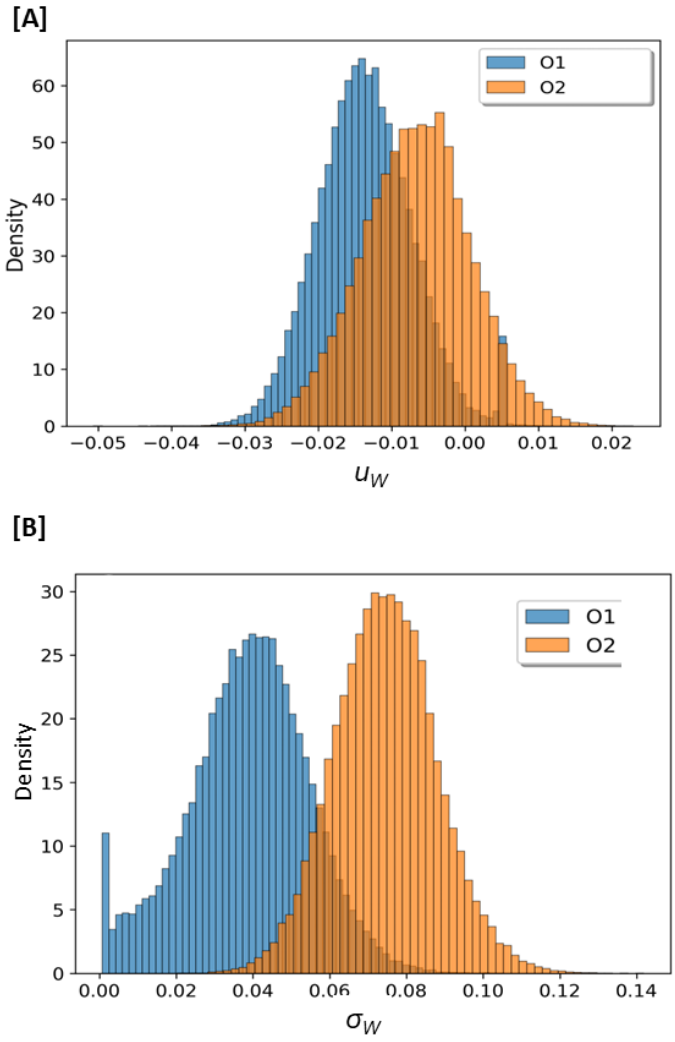
Alt Text: Histograms showing the distribution of simulated values of the difference between first-order and second-order mutation rates as a function of the number of false positive inferences. Panel (A) shows the distribution for simulated SNVs, panel (B) shows the distribution for simulated indels.

Figure 4. Type-specific mutation rates and spectra. (**Top left**) SNV mutation rates (**Top right**) SNV frequency spectrum **(Bottom left**) Indel mutation rates (**Bottom right**) Indel frequency spectrum. Error bars are SEM of MA lines.



Alt Text: Bar graphs showing the distributions of mutation rates and spectra. The top panels show SNVs, the bottom panels show indels; left panels show mutation rates, right panels show mutation spectra.

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



Alt Text: Histograms depicting the distribution of inferred mutational effects on competitive fitness. Panel (A) depicts mean effect, panel (B) depicts the standard deviation of the mutational effect.

Table 1. Summary statistics. 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 ($\times 10^9$) including correction for false inferences at the 3X coverage criterion. Estimated mutation rates ($\hat{\mu}$) are calculated as $\mu = \frac{m-FP+FtR}{L\times t}$, where *m* is the observed number of variants, *FP* is the number of false positives, *FtR* is the number of mutations failed to recall and *L* is the number of bases called at 3X coverage. See Methods for details of the calculations.

	N_{Lines}	***%Genome*** (SEM)	$\bar{t}$ (SEM)	$\bar{\mu}_{SNV}$ (SEM)	$\bar{\mu}_{DEL}$ (SEM)	$\bar{\mu}_{INS}$ (SEM)	$\bar{\mu}_{TOTAL}$ (SEM)	$\hat{\mu}_{SNV}$ (95 % CI)	$\hat{\mu}_{DEL}$ (95% CI)	$\hat{\mu}_{INS}$ (95% CI)	$\hat{\mu}_{TOTAL}$ (95% CI)
O 1	96	97.25 (0.051)	145.9 (0.8)	1.77 (0.027)	0.25 (0.011)	0.113 (0.0071)	2.13 (0.030)	1.78 (1.73-1.84)	0.25 (0.23--0.27)	0.12 (0.10--0.13)	2.15 (2.09--2.21)
			600	0.427 (0.0062)	0.061 (0.0027)	0.028 (0.0017)	0.52 (0.0072)	0.43 (0.42-0.44)	0.062 (0.056-0.067)	0.028 (0.025-0.032)	0.52 (0.508-0.536)
O 2	192	97.25 (0.051)	144.8 (1.6)	1.88 (0.032)	0.37 (0.016)	0.23 (0.013)	2.47 (0.046)	1.75 (1.68-	0.36	0.22	2.33

								1.81)	(0.33 - 0.39)	(0.20 - 0.25)	(2.24 - 2.42)
			624	0.43 (0.0075)	0.084 (0.0034)	0.053 (0.0026)	0.57 (0.0098)	0.40 (0.39-0.42)	0.082 (0.075-0.087)	0.051 (0.046-0.056)	0.54 (0.52 - 0.56)

1

2

3

4