Supplemental Methods

Table of Contents

1. Strains. *(i) MA lines*. The *mev-1* gene encodes a subunit of succinate dehydrogenase cytochrome b, a component of complex II of the mitochondrial electron transport chain. The *mev-1(kn1)* mutation is a G→A transition that replaces a glycine with a glutamate, resulting in decreased enzyme activity and increased electron leak. The resultant phenotype has been explored in several species (Ishii et al. 1990; Ishii et al. 2011; Ishii et al. 2013; Ishii et al. 2016) and includes elevated 8-oxodG and mutation (Hartman et al. 2004; Ishii et al. 2005). The *mev-1(kn1)* allele was backcrossed into the canonical Baer lab N2 strain genomic background for 12 generations, followed by six generations of selfing by single-worm transfer to render the introgressed strain homozygous and to allow it to reach mutation-drift equilibrium. Consistent with previous reports (Ishii et al. 1990), the introgressed *mev-1* ancestor had lower total reproductive output and shorter lifespan relative to N2 (Joyner-Matos et al. 2011; Dossey 2014). Importantly, our *mev-1* ancestor exhibited extreme sensitivity to methyl viologen (paraquat); in fact, the name "*mev-1*" is an abbreviation for "abnormal MEthyl Viologen sensitivity" (Ishii et al. 1990). In preliminary assays with 0.25 mM paraquat at 20°C, all (> 50) N2 worms survived and reproduced; none of the > 50 *mev-1* worms reached reproductive maturity, regardless of whether they were transferred to paraquat-containing plates as eggs or as L1 larvae (M. Dossey and JJ-M, unpublished results).

The sequence of the *mev-1* locus was inspected by eye in all *mev-1* MA lines. One line (735) did not have the *mev-1(kn1)* allele, presumably due to a cross-contamination event with an N2 strain, and was omitted from the analysis. All other *mev-1* lines were confirmed as homozygous for the *mev-1(kn1)* allele.

2. Mutation accumulation (MA) experiments.

Details of the N2 and PB306 mutation accumulation experiment are given in Baer et al. (2005), those of the *mev-1* MA experiment are given in Joyner-Matos et al. (2011). The basic protocol in both experiments follows that of Vassilieva and Lynch (1999) and is depicted in **Figure 1** in

the main text. Briefly, replicate populations (MA lines) were initiated from a cryopreserved stock of a highly-inbred ancestor at mutation-drift equilibrium and maintained over the course of the experiment by transferring a single immature hermaphrodite at four-day intervals. Worms were maintained on 60mm NGM agar plates, spotted with 100 *µl* of an overnight culture of the OP50 strain of *E. coli* B, at a constant 20°C.

At every generation, the prior two generations of each MA line were kept at 20° C as backups; if the leading generation worm did not reproduce, the plate was reinitiated with an immature individual from the previous generation; this is referred to as "going to backup". If the parental worm had reproduced but no offspring had reached the L2 stage, we kept the offspring in the experiment to reproduce in the next generation; we refer to this delay in reproduction as "holding over". Holding over reduces the actual number of generations of evolution of an MA line below the maximum (Gmax). Ideally, going to backup should not affect the total number of generations of MA, because the backup worm should be a (double) first cousin of the worm that would have been transferred. However, there is some opportunity for overlapping generations on backup plates (e.g., if we had to go to backup twice or more in a row), so there is some ambiguity with respect to the true number of generations experienced by a line. The upper bound on the number of generations is the total number of possible transfers in the experiment (Gmax); the lower bound is the number of successful transfers (Gmin). Data on transfers and additional context is presented in **Supplemental Table S1**.

3. DNA Extraction and library preparation. (i) MA lines. 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 (~50 mm^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 6× loading dye and run on a 2% agarose gel. DNA segments of 300-500bp 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 on a NovaSeq 6000 by Novogene, Inc.

(ii) Wild isolates. Details of the DNA extraction and library preparation are reported on the CeNDR website (https://www.elegansvariation.org/) and in Lee et al. (2021). Briefly, worms were cultured on two 10-cm NGMA plates spotted with OP50, washed into a 15-ml conical tube in M9 buffer and allowed to settle. Worms were rinsed 3× in 10 ml of M9 over 1 h to allow worms to purge ingested *E. coli*. Genomic DNA was isolated from 100–300 µl pellets using the Blood & Tissue DNA isolation kit (69506; Qiagen) following established protocols. DNA concentration of each sample was quantified using the Qubit dsDNA Broad Range Assay Kit (Q32850; Invitrogen). DNA samples were then submitted to the Duke Sequencing and Genomic Technologies Shared Resource per their requirements. The Illumina library construction and sequencing were performed at Duke University using KAPA HyperPrep kits (Kapa Biosystems) and sequenced on the Illumina NovaSeq 6000 platform (150 bp paired-end reads).

4. Variant calling. (i) MA lines. Adapter sequence was trimmed from raw sequencing reads using fastp (Chen et al. 2018). Following trimming, we used bowtie2 (Langmead et al. 2009) to align trimmed sequence data to the N2 reference genome (WS263). Reads with the MQ<3 were removed from further analyses. Duplicate reads were identified and removed using MarkDuplicates tool in Genome Analysis Toolkit (GATK)/picard. Variants (SNPs and indels) were called using HaplotypeCaller (in BP_RESOLUTION mode) in GATK4 (v4.1.4.0) (McKenna et al. 2010). Next, the resulting files from all samples (23 *mev-1*, 68 N2, and 67 PB306 samples plus their ancestors) were consolidated using GenomicsDBImport, and variants called jointly using GenotypeGVCFs in GATK.

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 MA line for each datastore, and (3) the ancestor genotype is homozygous wild-type (Saxena al., 2019). We initially applied a 3x coverage threshold for filtering the sites with <3x coverage and include only sites that are covered >3× in > 90% of the MA lines in each strain (for example 62 out of 68 in N2 MA lines). Putative indels ≥ 20 bp and all indels within 50 bp of another putative indel were visually inspected using the Integrative Genomics Viewer (IGV) software.

Potential functional effects of mutations were assessed using snpEFF 4.3 (Cingolani et al. 2012), and the annotation database WBcel235.82. Each variant potentially has multiple effects; snpEFF sorts them by potential impact (i.e., the highest putative impact is listed first). We include only the largest potential effect of each variant (**Supplemental Table S3**).

(ii) Wild isolates. The details of the CeNDR variant calling pipeline are given on the CeNDR website (https://www.elegansvariation.org/) and in Lee et al. (2021). In brief, adapters and low quality sequences were trimmed from raw reads using fastp and default parameters (https://github.com/andersenlab/trim-fq-nf). Trimmed reads were aligned to the *C. elegans* reference genome (WS276) using bwa mem BWA (0.7.17) (Li and Durbin 2010). Libraries of the same strain were merged together and indexed by sambamba (0.7.0). Duplicates were flagged

with Picard (2.21.3) (https://github.com/andersenlab/alignment-nf). Variants for each strain were called using gatk HaplotypeCaller. After the initial variant calling, variants were combined and then recalled jointly using gatk GenomicsDBImport and gatk GenotypeGVCFs GATK (4.1.4.0) (https://github.com/andersenlab/wi-gatk). We applied a ≥3x coverage threshold,and include only sites that are covered >3× in > 90% of the wild isolates.

The only difference between the CeNDR variant-calling pipeline and our MA mutation-calling pipeline is that the CeNDR pipeline uses BWA as the mapping tool whereas our mutation-calling pipeline uses Bowtie2. To address the possibility that the discrepancy in the mutation spectra between MA lines and wild isolates can be explained by the use of different mapping software, we re-analyzed the three sets of MA lines, *mev-1,* N2 and PB306 using the CeNDR pipeline (i.e., using BWA rather than Bowtie2 as the aligner). The resulting mutation spectra are nearly identical, both for base-substitutions and indels (**Supplemental Figures S6, S7**). Thus, we conclude that the discrepancy between the MA spectra and the wild isolate spectra cannot be explained by the use of different mapping software.

5. Data Analysis: General Linear Models (GLM). Line-specific mutation rates were used as the dependent variable in analyses of mutation rate. The model can be formally written as $y_{ijk} = \mu + a_j + b_k + ab_{ik} + e_{ijk}$, where y_{ijk} is the rate of mutation type j in line i of strain k, μ is the overall mean, a_j is the fixed effect of mutation type j, b_k is the fixed effect of strain k, ab_{bik} is the fixed effect of the interaction between mutation type j and strain k, and e_{ijk} is the residual effect associated with MA line i and where it is assumed that $e \sim N(0,\mathbf{R})$, with covariance structure **R** (see below). Covariances were estimated by restricted maximum likelihood (REML); fixed effects were tested by F-tests with Type III sums of squares. Analyses were implemented using the MIXED procedure in SAS v. 9.4.

We compared models with different covariance structures (**R**) by finding the model that minimizes the corrected Akaike Information Criterion (AICc). If the best model by AICc includes

147 more parameters than the next-best model, we compared the two by Likelihood Ratio Test
148 (LRT). The models are nested, so twice the difference in the log-likelihoods is asymptotically
149 chi-square distributed with degrees of freedom equal to the difference in the number of
150 parameters estimated. We initially pooled the data across strains to find the best model, then
151 compared that model to a model with the covariance structure estimated separately for each
152 strain and tested by the same criteria. The covariance structures tested, in increasing order of
153 complexity, are: variance components (a single, common variance and all covariances
154 constrained to equal zero), compound symmetry (common variance, common covariance);
155 banded main diagonal (a separate variance estimated for each variable, all covariances
156 constrained to zero), and unstructured (all elements of the covariance matrix estimated
157 individually).

158 Preliminary analyses revealed qualitatively different mutation rates and spectra for
159 mononucleotide repeats (defined as five or more consecutive bases of the same type)
160 compared to all other sequence motifs; see section 8 below for details of the bioinformatics by
161 which mononucleotide repeats were identified. Rather than extend the above GLM to include
162 sequence type (mononucleotide and non-mononucleotide), which would have included three
163 two-way interactions as well as the three-way interaction, we broke the analysis down into two
164 different two-factor models. First, we analyzed each of the six base-substitution rates
165 individually, including strain and sequence type as predictor variables. Comparisons among
166 strains were restricted to the comparison between *mev-1* and N2 (the foundational motivation
167 for the study) and the comparison between N2 and PB306. Next, since in only one case did the
168 strain-by-sequence type interaction approach statistical significance (see Results), we pooled
169 the data over strains (N2 and PB306, omitting *mev-1*) and analyzed the model with sequence
170 type and mutation type as the predictor variables.
171

6. mtDNA mutation rate. Estimation of the mtDNA mutation rate is more complicated than for nuclear loci because a non-trivial fraction of mutations will not have reached fixation and remain heteroplasmic. The probability of a mutation ultimately reaching fixation is its current frequency in the population (Wright 1931), so the mtDNA mutation rate of MA line *i* can be estimated as $\mu_i = \frac{\sum p_j}{n_i t_i}$, where p_j is the frequency of the *j*'th mutation in line *i*, n_i is the number of callable sites in line *i*, and t_i is the number of generations of MA of line *i* (Konrad et al. 2017).

Because many lines have no mtDNA mutations, the residuals of mtDNA mutation rate are far from normally distributed, so GLM is not an appropriate method by which to test the hypothesis that mtDNA mutation rates differ among strains. Instead, we used a randomization test, in which (1) the number of lines/strain is held constant and the data (line-specific mutation rates) randomly shuffled among strains, (2) the average mutation rate calculated for each strain, and (3) the difference in average mutation rate between the three pairs of strains calculated. This procedure was repeated 1000 times, to generate a null distribution of pairwise differences in mutation rate. If the observed value of the pairwise difference is greater than 95% of the randomized values, we take it to be statistically significant at $P<0.05$.

7. Parametric bootstrap simulations of wild isolate mutation spectra. To test the hypothesis that the variance in the private allele frequency spectrum among wild isolates can be explained by sampling variance around a single uniform base-substitution spectrum with expectation equal to the observed frequencies, we employed parametric bootstrap simulations, as follows. First, each wild isolate was assigned its observed number of private alleles, sampled with replacement with probability equal to the observed frequency (e.g., if the frequency of A:T→T:A base substitutions is 20%, the probability that any simulated private allele is the result of a A:T→T:A mutation is 20%). Next, the Kullback-Leibler divergence (Kullback and Leibler 1951), $D_{KL} = \sum_{i=1}^{n} p_i log \frac{p_i}{q_i}$, was calculated for each wild isolate, where p_i and q_i are the observed and

expected base-substitution frequencies, and the mean divergence $\overline{D}_{KL}$ of the set of wild isolates calculated, as well as the variance in D_{KL} among wild isolates. This procedure was repeated 1000 times to generate a simulated frequency distribution of the mean D_{KL} and its variance under the null hypothesis, and the observed values compared to the null distribution. If the observed value is greater than 95% of simulated values, we reject the null hypothesis at the $p<0.05$ level.

8. Mononucleotide repeats. We used Phobos 3.3.12 software (http://www.rub.de/ecoevo/cm/cm_phobos.htm) to identify all mononucleotide repeats of ≥ 5 bp in the WS263 build of the *C. elegans* genome. To identify imperfect repeats, we applied the default mismatch penalties (mismatch score -5, gap score -5, recursion depth 5, Minimum score = 0). The minimum repeat unit criterion was 5bp and we kept all the sites with 100% repeat perfection. We increased the range of all mononucleotide repeat by two bases on either end of the repeat, following Saxena et al. (2019). The average coverage of mononucleotide repeats and non-mononucleotide repeat in the MA lines of the three strains was estimated using mosdepth 0.3.1 (Pedersen and Quinlan 2018); the same coverage threshold was applied for mononucleotide and non-mononucleotide regions. Including the 2 bp flanking regions, mononucleotide repeats include 14,470,735 nucleotides and non-mononucleotide regions include 85,815,671 nucleotides. Excluding the flanking regions, mononucleotide repeats *per se* include 8,578,075 nucleotides. The GC-content of mononucleotide and non-mononucleotide sequence is 20.12% and 38.02%, respectively.

9. Local sequence context. We counted the total number of each of the 64 3-base motifs in the reference genome using Jellyfish v. 2.3.0 software (Marcais and Kingsford 2011). For example, in the sequence -5'-AAAA-3', there are two possible 5'-aAa-3' motifs. We then applied a custom

Bash script to count the number of each motif in the MA lines. We used bcftools to extract the position of each variant, and bedtools to extract the sequence context of each variant.

We counted the total number of mutations occurring at the middle position in each of the 3-bp motifs and estimated the base-substitution mutation rate of each 3-base motif *i* in an MA line as $\mu_i = \frac{X_i}{n_i t}$, where *X* is the total number of mutations in each 3-bp motif in the MA line, *n* is the detectable number of each 3-base motif in that line, and *t* is the number of generations of MA. The reported values of μ_i are unweighted means over all lines in the strain. Scripts are deposited in github at https://github.com/moeinrajaei/mev-1-project.

10. False positives (FP) and false negatives (FN).

(i) False Positives. False positives were assessed in two ways. First, we counted sites that differ between our MA ancestor and the reference (N2) genome (i.e., sites that are scored 1/1 in the MA ancestor and 0/0 in the reference genome). These are presumed to be new mutations fixed subsequent to the divergence of the MA ancestor and the reference strain from their common ancestor. These sites were scored in each MA line in the set. If a site is scored 1/1 in the MA ancestor but scored 0/0 in all MA lines, we inferred a false positive in the MA ancestor. The power of this method is obviously sensitive to the number of MA lines sequenced; the more MA lines sequenced, the greater the chance of identifying a false positive. Since the number of N2 lines (68) and PB306 lines (67) are nearly equal, any difference in the frequency of false positives cannot be attributed to differences in power.

Second, we employed an independent set of low-coverage sequence data from a set of 192 recombinant inbred advanced intercross lines (RIAILs) generated from a cross between two N2 strain MA lines, 530 and 563. The details of the construction and sequencing of the RIAILs and data analysis are presented insection 12 below. .

(ii) False negatives. To identify false negatives, we first introduced simulated variants ("dummy mutations") into the reference genome, and then re-analyzed the MA data using the simulated reference genome. For example, suppose position *x* in the reference genome is A. We randomly change the A to a C, i.e. an A →C "dummy" transversion. In this case, the MA ancestor and all MA lines should have a C at site *x*. If the genotype at site *x* in any MA line is not called a C/C homozygote, for any reason, it is classified as a failure to recall. If the genotype at site *x* in an MA line was called "not C/C" (i.e., it was scored A/A or A/C), it is a false negative. We introduced 983 dummy base substitutions into the reference genome; failure to recall rates are given in the main text.

Indels present a more complex problem, because inserting or deleting sequence changes the physical map of the reference genome relative to the focal genome. To address the question of indel false negatives, we first introduced 9,990 dummy indels into the reference genome (**Supplemental Table S12** below). Next, we defined five size bins of 1-5 bp, 6-10 bp, 11-20bp, 21-50bp, and >50 bp, and analyzed the data as described above for base-substitutions. For each MA line, we counted the total number of indels in each bin using the simulated reference genome and the true reference genome and compared the number of indels in each bin. We then repeated the simulation and analyses with a second set of dummy mutations; the correlation between the line-specific failure to recall rates in the two simulations provides an estimate of the reliability of the method (**Supplemental Figure S12** below).

11. Indel filters

To compare the indel spectra of MA lines and segregating private alleles, we applied the following filtering criteria using VariantFiltration in GATK (4.1.4.0). Variants with quality (QUAL) < 30, quality by depth (QD) < 20, read depth (DP) < 5, strand odds ratio (SOR) > 5, Fisherstrand (FS) > 100, ReadPosRankSum value < -20. Sites with > 10% missing genotypes and or any heterozygous genotypes were removed.

12. Recombinant Inbred Advanced Intercross Lines (RIAILs)

12.1. Line Construction. Two N2-strain MA lines, line 530 (243 MA generations) and line 563 (237 MA generations) were chosen as parents to generate a panel of 200 recombinant inbred advanced intercross lines (RIAILs) (Rockman and Kruglyak 2008). In brief, we used a random pair mating design with equal contributions of each parent to the next generation over ten generations of intercrossing, followed by ten generations of selfing. We first thawed the cryopreserved parental strains, lines 530 and 563, on 100mm plates seeded with OP50 *E. coli*. Both strains contained males when thawed, and we allowed them to propagate until many gravid hermaphrodites were present. We then bleached the strains and incubated the recovered embryos in M9 buffer overnight to make synchronized cultures of arrested L1 stage larvae. We plated these L1s on multiple 100 mm plates to avoid overcrowding and grew them for 48 hours. We then set up 30 independent outcrosses on mating plates (30mm plates seeded with 10uL of OP50 *E. coli*) with the sex of the parental strains reversed for half the outcrosses. We set up these and all subsequent crosses with two young males and one virgin hermaphrodite each, and always transferred the parents to a new mating plate two days after setting up the cross to avoid accidentally selecting self progeny. We made 280 independent crosses among F1 progeny, being careful to select animals that would provide equal proportions of the founders' cytoplasmic and X chromosomes to the F2 progeny. Of these crosses, only 200 showed the roughly equal sex ratios in the progeny indicative of successful mating, therefore all subsequent crosses were made in duplicate to avoid failed matings in subsequent generations. We randomly paired F2 progeny from the 200 successful F1 crosses to generate the F3 generation and performed each cross reciprocally to maintain a constant population size of 200. We repeated this process over ten generations of intercrossing. In some cases, both duplicate crosses failed, in which case we randomly selected another line of the same generation to take its place, leading to the occasional overrepresentation of some parents within a generation of

intercrossing. At the F10 generation we chose a single hermaphrodite from each of the 200 recombinant lines and allowed them to self fertilize for 10 generations. Of the 200 RIAILs, 196 survived cryopreservation and were used in this study.

12.2. Illumina library construction and whole-genome sequencing. To isolate genomic DNA from the recombinants, we transferred nematodes from two freshly starved 10 cm NGM plates into a 15 ml conical tube by washing with 10 mL of M9 buffer. We allowed the worms to settle by gravity the bottom of the conical tube, removed the supernatant, and added 10 mL of fresh M9. We repeated this wash method three times over the course of one hour to serially dilute any remaining *E. coli* in the M9 and allow the nematodes time to purge ingested bacteria. We then isolated genomic DNA from 100 to 300 µl nematode pellets using the Blood and Tissue DNA isolation kit cat# 69506 (QIAGEN, Valencia, CA) following established protocols (Cook et al. 2016). We quantified the concentration of DNA in each sample with the Qubit dsDNA Broad Range Assay Kit cat# Q32850 (Invitrogen, Carlsbad, CA) and diluted all samples to 0.2 ng/µL. We then incubated each sample with diluted Illumina transposome cat# FC-121-1031 (Illumina, San Diego, CA) and amplified the tagmented fragments with barcoded primers. We combined the uniquely indexed samples into 96-sample pooled libraries and size-selected them by separating the material on a 2% agarose gel, extracting the fragments ranging from 400-600 bp, and purifying the extraction with the QIAquick Gel Extraction Kit cat# 28706 (QIAGEN, Valencia, CA). We determined the concentration and size distribution of the purified, size-selected libraries with the 2100 Bioanalyzer (Agilent, Santa Clara, CA) before submitting them to the NUSeq Core Facility of Northwestern University where they were sequenced on the Illumina HiSeq4000 platform (paired-end 150 bp reads).

12.3. Variant calling and analysis. Variants were called using the same analysis pipeline as for the MA lines, described in section 4 above. We identified 699 SNPs and 1054 indels in the two parental lines (Note: these variants include new mutations in the MA phase and fixed mutations that occurred on the lineage leading to the MA ancestor subsequent to its divergence from its

common ancestor with the reference genome). The variants were recalled jointly in 192 RIAILs and 636 SNPs and 961 indels were kept. The expected frequency of a variant in the set of RIAILs is 50% (i.e., we expect Mendelian segregation). If a variant is not present in at least ten (out of 192) RIAILs and there are fewer than 50 RIAILs with missing data at the site, it is considered a false positive.

13. Note S1. Comparison of variant-calling pipelines on a Comparison Subset of lines.

The discrepancy in the mutation spectra between MA lines and wild isolates, especially in mononucleotide repeats, could potentially have technical as well as biological causes. One potential source of technical discrepancy is that, both within and between data sets, Illumina libraries were prepared at different times, using different protocols. Samples were sequenced on different sequencing platforms, with different paired-end read lengths, and with different average read depths. For one example, some samples were sequenced with 100 bp paired-end reads and others with 150 bp paired-end reads. Longer mononucleotide runs will be more difficult to identify in shorter reads, and will potentially be underrepresented in samples with 100 bp PEs relative to samples sequenced with 150 bp PEs.

We reasoned that, if the discrepancy between the MA spectra and the wild isolate spectra are biological rather than technical, comparison of MA and wild isolate samples that were prepared and sequenced in the same way would still differ. To that end, we chose a sample of 76 MA lines (38 N2 and 38 PB306) and a sample of 38 wild isolates for which the libraries were prepared at approximately the same time, in the same lab, using the same reagents and protocols, and were sequenced using 150 bp paired-end reads at approximately the same time on the same sequencing platform (NovaSeq 6000) by the same contractor (Novogene, Inc), to approximately the same depth of coverage. The libraries were prepared in the Andersen lab in 2019; the DNA extraction and library construction protocols are described in section 3 of the Supplemental Methods. Thus, to the best of our ability, we are comparing

apples to apples. We refer to this subset of lines as the "Comparison Subset". The list of lines and the details of sequencing are given in **Supplemental Table S11.**

We did not have access to the raw, unfiltered sequencing reads of the wild isolates; we began the analysis with the BAM files available on the CeNDR website (https://www.elegansvariation.org/) and we applied the same pipeline for variant calling for all three datasets – N2 and PB306 MA lines and the wild isolates. The resulting mutation spectra were very similar to the spectra reported in the main text (Figures 4 and 5), as depicted in **Supplemental Figures S4** (base-substitutions) and **S5** (indels).

Next, we applied a partial “reciprocal transplant” strategy, as outlined in section 4 of the Supplemental Methods. We analyzed the subset of MA lines described above using both our MA analysis pipeline, as described in section 4(i) of the Supplemental Methods, and the CeNDR bioinformatics pipeline outlined in section 4(ii) of the Supplemental Methods. The resulting mutation spectra are very similar in the two analyses, as depicted in **Supplemental Figure S6** (base-substitutions) and **S7** (indels).

Another possible source of the discrepancy in the mutation spectra, again particularly for mononucleotide repeats, is that our minimum coverage threshold for inclusion is ≥3×, which is low. We chose that low threshold because the number of false negatives (and thus the failure to recall rate) declines faster with increasing coverage than does the rate of false positives, so beyond a very low threshold, for the purposes of estimation of mutation rate, it is better to include more sites at low coverage than fewer sites at higher coverage. As an empirical test of the hypothesis that the discrepancy could be due to a high rate of false positives at low coverage, we repeated the analysis of the MA lines with a more stringent criterion of ≥10× coverage. Again, the spectra are very similar, both for the MA lines and the wild isolates (**Supplemental Figure S8 and S9 (base-subs**) and **S10 and S11** (indels)).

The conclusion from these additional analyses is simple: the two different methods of analysis (ours and the CeNDR pipeline) give concordant results on the same data, chosen for

376 methodological concordance. Moreover, increasing the stringency of the coverage threshold
377 from 3× to 10× does not change the results.
378

Literature Cited

Baer CF, Shaw F, Steding C, Baumgartner M, Hawkins A, Houppert A, Mason N, Reed M, Simonelic K, Woodard W et al. 2005. Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proceedings of the National Academy of Sciences of the United States of America* **102**(16): 5785-5790.

Chen SF, Zhou YQ, Chen YR, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17): 884-890.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly* **6**(2): 80-92.

Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga JE et al. 2016. The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics* **204**(1): 371-383.

Dossey MM. 2014. An exploration of mutation effect sizes in the nematode *Caenorhabditis elegans*. In *Department of Biology*, Vol M.S., p. 80. Eastern Washington University, Cheney, WA USA.

Hartman P, Ponder R, Lo HH, Ishii N. 2004. Mitochondrial oxidative stress can lead to nuclear hypermutability. *Mechanisms of Ageing and Development* **125**(6): 417-420.

Ishii N, Takahashi K, Tomita S, Keino T, Honda S, Yoshino K, Suzuki K. 1990. A methyl viologen sensitive mutant of the nematode *Caenorhabditis elegans*. *Mutation Research* **237**(3-4): 165-171.

Ishii T, Miyazawa M, Hartman PS, Ishii N. 2011. MItochondrial superoxide anion ($O2^{\cdot-}$) inducible "*mev-1*" animal models for aging research. *BMB Reports* **44**(5): 298-305.

Ishii T, Miyazawa M, Onouchi H, Yasuda K, Hartman PS, Ishii N. 2013. Model animals for the study of oxidative stress from complex II. *Biochim Biophys Acta* **1827**: 588-597.

Ishii T, Yasuda K, Akatsuka A, Hino O, Hartman PS, Ishii N. 2005. A mutation in the *SDHC* gene of Complex II increases oxidative stress, resulting in apoptosis and tumorigenesis. *Cancer Res* **65**: 203-209.

Ishii T, Yasuda K, Miyazawa M, Mitsushita J, Johnson TE, Hartman PS, Ishii N. 2016. Infertility and recurrent miscarriage with complex II deficiency-dependent mitochondrial oxidative stress in animal models. *Mech Ageing Dev* **155**: 22-35.

Joyner-Matos J, Bean LC, Richardson HL, Sammeli T, Baer CF. 2011. No evidence of elevated germline mutation acccumulation under oxidative stress in *Caenorhabditis elegans*. *Genetics* **189**(4): 1439-1447.

Konrad A, Thompson O, Waterston RH, Moerman DG, Keightley PD, Bergthorsson U, Katju V. 2017. Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in Caenorhabditis elegans Spontaneous Mutation Accumulation Lines of Differing Population Size. *Molecular Biology and Evolution* **34**(6): 1319-1334.

Kullback S, Leibler RA. 1951. On information and sufficiency. *Ann Math Statist* **22**(1): 79-86.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**(3).

Lee D, Zdraljevic S, Stevens L, Wang Y, Tanny RE, Crombie TA, Cook DE, Webster AK, Chirakar R, Baugh LR et al. 2021. Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nature Ecology & Evolution*.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**(5): 589-595.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**(6): 764-770.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**(9): 1297-1303.

Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**(5): 867-868.

Rockman MV, Kruglyak L. 2008. Breeding designs for recombinant inbred advanced intercross lines. *Genetics* **179**(2): 1069-1078.

Saxena AS, Salomon MP, Matsuba C, Yeh SD, Baer CF. 2019. Evolution of the mutational process under relaxed selection in *Caenorhabditis elegans*. *Molecular Biology and Evolution* **36**(2): 239-251.

Vassilieva LL, Lynch M. 1999. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151**(1): 119-129.

Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16**(2): 0097-0159.