Mutability of mononucleotide repeats, not oxidative stress, explains the discrepancy between laboratory-accumulated mutations and the natural allele-frequency spectrum in *C. elegans*

Moein Rajaei,[1] Ayush Shekhar Saxena,[1] Lindsay M. Johnson,[1] Michael C. Snyder,[1] Timothy A. Crombie,[1,2] Robyn E. Tanny,[2] Erik C. Andersen,[2] Joanna Joyner-Matos,[3] and Charles F. Baer[1,4]

[1]*Department of Biology, University of Florida, Gainesville, Florida 32611, USA;* [2]*Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, USA;* [3]*Department of Biology, Eastern Washington University, Cheney, Washington 99004, USA;* [4]*University of Florida Genetics Institute, Gainesville, Florida 32608, USA*

Important clues about natural selection can be gleaned from discrepancies between the properties of segregating genetic variants and of mutations accumulated experimentally under minimal selection, provided the mutational process is the same in the laboratory as in nature. The base-substitution spectrum differs between *C. elegans* laboratory mutation accumulation (MA) experiments and the standing site-frequency spectrum, which has been argued to be in part owing to increased oxidative stress in the laboratory environment. Using genome sequence data from *C. elegans* MA lines carrying a mutation (*mev-1*) that increases the cellular titer of reactive oxygen species (ROS), leading to increased oxidative stress, we find the base-substitution spectrum is similar between *mev-1*, its wild-type progenitor (N2), and another set of MA lines derived from a different wild strain (PB306). Conversely, the rate of short insertions is greater in *mev-1*, consistent with studies in other organisms in which environmental stress increased the rate of insertion–deletion mutations. Further, the mutational properties of mononucleotide repeats in all strains are different from those of nonmononucleotide sequence, both for indels and base-substitutions, and whereas the nonmononucleotide spectra are fairly similar between MA lines and wild isolates, the mononucleotide spectra are very different, with a greater frequency of A:T → T:A transversions and an increased proportion of ±1-bp indels. The discrepancy in mutational spectra between laboratory MA experiments and natural variation is likely owing to a consistent (but unknown) effect of the laboratory environment that manifests itself via different modes of mutability and/or repair at mononucleotide loci.

[Supplemental material is available for this article.]

It is a fundamental principle of population genetics that the DNA sequence diversity in a population (θ) represents the product of mutation (μ) and "everything else," where "everything else" subsumes the contributions of natural selection and random genetic drift in the composite parameter N_e, the genetic effective population size: $\theta = 4N_e\mu$ (Watterson 1975; Nei and Li 1979). Empirically partitioning genetic variation into the contributions of mutation and everything else requires that mutations be observed under conditions in which the effects of natural selection are minimized as much as possible. That can be performed in three basic ways: by genotyping related individuals in a natural pedigree, as is now performed commonly in humans (Gao et al. 2019; Halldorsson et al. 2019); by means of a "mutation accumulation" (MA) experiment; or by cataloging very rare segregating variants that have (presumably) arisen only recently and thus been minimally sieved by natural selection (Messer 2009; Carlson et al. 2018).

An MA experiment is in essence a large pedigree in which many replicate descendant lines (MA lines) are derived simultaneously from a common ancestor and allowed to evolve under conditions in which selection is minimal (Halligan and Keightley 2009). Typically, selection in an MA experiment is minimized by minimizing N_e, ideally to a single individual or chromosome. Pedigree genotyping has two important advantages over MA experiments: It can be performed in practically any organism, and it is fast. MA experiments are slow, labor intensive, and only practical in organisms with fast generation times that can be maintained easily in the laboratory. However, MA experiments come with one unique advantage, which is that the phenotypic effects of mutations can be assessed under controlled experimental conditions. MA experiments have been a workhorse of evolutionary genetics for the past 60 years (Sprague et al. 1960; Mukai 1964; Liu and Zhang 2019), and much of our understanding of the mutational process has been derived from MA data (Drake et al. 1998; Halligan and Keightley 2009; Katju and Bergthorsson 2019).

The utility of MA experiments is based on a key assumption: The mutational process under laboratory conditions faithfully

Corresponding authors: moeinraja@ufl.edu, cbaer@ufl.edu
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.275372.121.

31:1602–1613 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/21; www.genome.org

reflects that in nature. It has long been known from studies with microbes that all elements of the mutational process—rate, molecular spectrum, and phenotypic effects—depend on the environmental context and the genetic background; this conclusion has recently been extended to multicellular eukaryotes (Sharp and Agrawal 2012, 2016; Kessler et al. 2020). On the one hand, the context dependence of the mutational process is not a surprise; for example, the mutagenic effects of X-rays have been known for a century. On the other hand, it suggests that MA experiments come with their own biases (Baer 2019), just as natural selection biases the standing genetic variation.

The nematode *Caenorhabditis elegans* is an important model organism in many areas of biology and was the first metazoan organism to have its mutational process characterized at the genomic level (Denver et al. 2000, 2004b). As genomic sequence data from *C. elegans* MA lines and wild isolates have accumulated, it has become apparent that the base-substitution spectrum in laboratory-accumulated mutations differs from that of wild isolates in a consistent way: There are more transversions in the laboratory than there are in nature. The ratio of transitions to transversions (Ts/Tv) in MA lines is consistently around 0.6–0.8, whereas the Ts/Tv ratio among wild isolates is around 1.1–1.2. Two early studies reported that G:C → T:A transversions are overrepresented in MA data relative to the standing nucleotide diversity (Denver et al. 2009, 2012); two more recent studies with more data reported that A:T → T:A transversions are also overrepresented in MA data relative to the standing nucleotide diversity (Konrad et al. 2019; Saxena et al. 2019).

The MA spectrum may differ from the standing spectrum for several reasons. First, it may be that the mutational milieu in the laboratory is consistently different from that in nature. Second, it may be that purifying selection against transversions in nature is stronger than against transitions, and there is some reason to think that idea is plausible (Babbitt and Cotter 2011; Guo et al. 2017). Third, it may be that the analyses of MA data and standing genetic variation come with different biases.

One key difference between the laboratory MA environment and the natural environment is that worms in the laboratory are fed ad libitum and kept at low population density and at a constant benign temperature, which may result in a long-term elevation of metabolic rate relative to that experienced in nature. A second key difference is that worms kept on plates in the laboratory experience near-atmospheric concentrations of O_2, which is substantially greater than the optimum, as assayed by worm preference (Gray et al. 2004). Both increased metabolic rate and increased O_2 partial pressure could potentially increase the cellular concentrations of reactive oxygen species (ROS). ROS are present in all cells as a natural byproduct of cellular metabolism and can induce potentially mutagenic oxidative damage to DNA (Cooke et al. 2003; Bridge et al. 2014) and alterations in chromatin structure (for review, see Kreuz and Fischle 2016). Accordingly, variation in cellular ROS levels has long been invoked as a potential cause of variation in mutation rate (Richter et al. 1988; Shigenaga et al. 1989; Martin and Palumbi 1993; Stoltzfus 2008). Cellular oxidative stress can vary because of differences in input (e.g., ROS levels increase under conditions of physiological stress), in the ability of cells to enzymatically convert ROS into benign products, and in repair processes (Beckman and Ames 1998; Turrens 2003; Halliwell and Gutteridge 2007; Constantini 2014).

One well-documented manifestation of oxidative damage to DNA is the oxidation of guanine to 8-oxo-7,8-dihydroguanine (8-oxodG), which if unrepaired results in a G → T transversion (Cheng et al. 1992; Cunningham 1997; Helbock et al. 1998; Cadet et al. 2003; Evans and Cooke 2004). Accordingly, G:C → T:A transversions are often interpreted as a signature of oxidative damage to DNA (Krašovec et al. 2017; Suzuki and Kamiya 2017; Poetsch et al. 2018). Although much of the early work linking ROS with 8-oxodG and mutation was conducted in somatic cells (Dollé et al. 2000; Busuttil et al. 2003), this work has been extended to sperm (e.g., Paul et al. 2011; Kim and Velando 2020). Because previous studies have found that the frequency of G:C → T:A transversions is greater in MA lines than in the standing variation, we were led to speculate that some property of the laboratory MA environment increases oxidative stress relative to that experienced in nature. To test that hypothesis, we used a mutant strain of *C. elegans*, *mev-1*, that experiences elevated steady-state oxidative stress (Ishii et al. 1998, 2013; Senoo-Matsuda et al. 2001). We reasoned that if G:C → T:A transversions are in fact a signature of oxidative damage to DNA, then a strain with an endogenously increased oxidative damage should show an even greater frequency of G:C → T:A transversions than the background MA frequency.

Here we report results from an experiment in which a set of MA lines derived from a strain carrying the *mev-1* (previously known as *kn1*) mutation backcrossed into our canonical wild-type N2 background was propagated for about 125 generations using our standard *C. elegans* MA protocol (Joyner-Matos et al. 2011). We sequenced the genomes of 23 *mev-1* lines and compared the rate and spectrum of mutation to that of the N2 strain. As a further comparison, we report results from an additional set of 67 MA lines derived from a different wild-type strain, PB306. The laboratory MA spectrum is compared with the natural mutation spectrum inferred from "private alleles" present as homozygous variants in one, and only one, wild isolate ($n = 773$ wild isolates).

Results

Mutation rate

Mononucleotide repeats mutate differently than other sequence. We first report genome-wide rates, which are ultimately of the most evolutionary relevance; we parse mutation rates into mononucleotide and nonmononucleotide rates in the next section. The experimental design is depicted in Figure 1.

Genome-wide mutation rate

Summary statistics on nuclear mutation rates are presented in Table 1 and results of statistical tests in Supplemental Table S2. Results for individual MA lines are given in Supplemental Table S3. A complete list of mutations and their genomic context is given in Supplemental Table S4 (nuclear loci) and Supplemental Table S5 (mtDNA loci).

To test the hypothesis that the overall base-substitution mutation rate μ_{BS} differs between strains, we used a general linear model (GLM); details are presented in the Methods. The hypothesis that the rate of G:C → T:A transversions is greater in *mev-1* than in N2 is a prior one-tailed hypothesis test. The GLM revealed significant variation in μ_{BS} among the three strains, as well as a significant interaction between strain and mutation type. Contrary to our expectation, μ_{BS} summed over all six mutation types is significantly less in the *mev-1* lines than in N2 and is marginally lower than in PB306. In contrast, μ_{BS} does not differ between N2 and PB306. The G:C → T:A transversion rate differs among strains (Fig. 2A), but not in the predicted way: $\mu_{GC \to TA}$ is indistinguishable between *mev-1* and N2 and lower is than in PB306. The rate of A:T

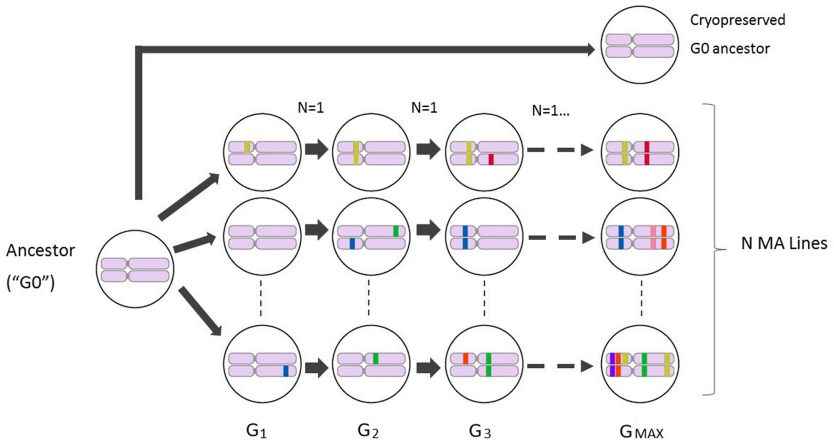

Figure 1. Propagation of MA lines. The common ancestor of the MA lines (G0) was thawed from a cryopreserved sample and a single immature hermaphrodite picked onto an agar plate. Lines were propagated by single-worm transfer at 4-d (one generation) intervals for t = Gmax transfers. Lines were initially genetically homogeneous (pink chromosome pairs); colored bars represent new mutations, which are fixed in a line with expectation $u=(1-s)/(2-s)$, where s is the selection coefficient (Keightley and Caballero 1997). For details of the MA experiments, see Methods and Supplemental Table S1.

→T:A transversions is marginally different between the three strains (Fig. 2A), with N2 having a slightly greater rate than *mev-1* and PB306.

The insertion and deletion rates both vary among strains, in different ways (Fig. 2D; Table 1). The PB306 deletion rate is significantly greater than that of *mev-1* and N2. The *mev-1* insertion rate is significantly greater than that of N2 and is greater than that of PB306, although not significantly. The PB306 insertion rate is marginally greater than that of N2.

The mechanisms of mutagenesis and repair are not necessarily uncorrelated, and the large number of mutations (more than 9000) provides an opportunity to investigate the covariance structure of the different types of mutation. We first investigated the covariance structure of the six types of base-substitution. We initially fit a model with a single variance component, that is, a uniform diagonal element in the covariance matrix and all off-diagonal elements constrained to zero. That model was compared to a model in which a separate variance (diagonal element) was estimated for each of the six traits ("banded main diagonal") and the off-diagonal elements constrained to zero. That model was in turn compared with a model in which all elements are unconstrained. The unconstrained model had the smallest AICc (ΔAICc = 16.2 less than the banded main diagonal) and provided a significantly better fit to the data (LRT, $-2\Delta\ln L = 47.1$, df = 15, $P < 0.0001$), although the individual correlations were small (all $|r| < 0.3$) (Supplemental Table S6). A model with the covariance matrix estimated separately for each strain did not fit as well (ΔAICc = 12.4 greater than the model with a single covariance matrix).

We next investigated the covariance structure of indels and base-substitutions together, combining all six base-substitutions into one category with insertions and deletions included separately. The best model is the most parameter-rich, with the unstructured covariance matrix estimated separately for each strain (ΔAICc = 12.3 less than the model with a single unstructured covariance matrix; LRT, $-2\Delta\ln L = 37.6$, df = 12, $P < 0.0002$). In all three strains, correlations between base-substitution and deletion rates are moderately positive ($r = 0.2$–0.3), and base-substitution and insertion rates are uncorrelated ($r \approx 0$). The correlation between base-substitution rate and insertion rate is negative in *mev-1* and N2, whereas it is moderately positive in PB306 ($r \approx 0.3$) (Supplemental Table S7). We emphasize that, for both of these covariance analyses, we cannot reject the hypothesis that any specific off-diagonal element is different from zero. Rather, the results simply mean that models with some nonzero off-diagonal elements fit the data better than a model in which all off-diagonal elements are constrained to zero, and we interpret the point estimates of the individual pairwise correlations heuristically.

Inspection of the residuals of the GLM reveals two MA lines that are obvious outliers. Line 572 (N2) has anomalously high base-substitution and deletion rates, although its insertion rate is slightly below the N2 average. Line 481 (PB306) has anomalously high insertion and deletion rates, although its base-substitution rate is only slightly greater than the PB306 average (Supplemental Table S3). We searched the list of mutations in those two lines for

Table 1. Mutation rates (μ) × 10^9/site/generation (SEM)

	mev-1 (n = 23 MA lines)			N2 (n = 68 MA lines)			PB306 (n = 67 MA lines)		
	Nonmono	Mono	Total	Nonmono	Mono	Total	Nonmono	Mono	Total
$\mu_{GC\rightarrow AT}$	1.45 (0.12)	1.42 (0.36)	1.45 (0.10)	1.83 (0.06)	1.6 (0.24)	1.81 (0.06)	1.81 (0.05)	1.91 (0.22)	1.82 (0.05)
$\mu_{GC\rightarrow TA}$	0.81 (0.12)	1.29 (0.36)	0.85 (0.11)	0.80 (0.04)	1.13 (0.18)	0.82 (0.04)	0.94 (0.04)	1.68 (0.21)	1.00 (0.05)
$\mu_{GC\rightarrow CG}$	0.41 (0.06)	0.65 (0.26)	0.43 (0.06)	0.46 (0.04)	0.42 (0.09)	0.46 (0.03)	0.43 (0.03)	0.42 (0.09)	0.43 (0.03)
$\mu_{AT\rightarrow GC}$	0.39 (0.05)	0.33 (0.12)	0.38 (0.04)	0.47 (0.03)	0.40 (0.05)	0.46 (0.02)	0.45 (0.03)	0.34 (0.04)	0.43 (0.02)
$\mu_{AT\rightarrow CG}$	0.20 (0.03)	0.32 (0.10)	0.22 (0.03)	0.29 (0.02)	0.39 (0.05)	0.31 (0.02)	0.25 (0.02)	0.35 (0.04)	0.27 (0.02)
$\mu_{AT\rightarrow TA}$	0.35 (0.05)	3.04 (0.29)	0.83 (0.06)	0.44 (0.03)	3.34 (0.16)	0.95 (0.04)	0.40 (0.02)	2.70 (0.14)	0.81 (0.03)
μ_{BS}	1.60 (0.07)	3.62 (0.26)	1.89 (0.07)	1.91 (0.05)	3.93 (0.15)	2.20 (0.05)	1.89 (0.05)	3.52 (0.16)	2.13 (0.05)
μ_{INS}	0.15 (0.03)	1.32 (0.19)	0.32 (0.04)	0.07 (0.01)	1.07 (0.06)	0.21 (0.01)	0.10 (0.01)	1.19 (0.09)	0.26 (0.02)
μ_{DEL}	0.15 (0.02)	1.76 (0.23)	0.38 (0.03)	0.19 (0.01)	1.43 (0.09)	0.36 (0.02)	0.26 (0.02)	1.76 (0.12)	0.47 (0.03)
μ_{INDEL}	0.30 (0.05)	3.09 (0.22)	0.70 (0.04)	0.25 (0.01)	2.50 (0.11)	0.58 (0.02)	0.36 (0.02)	2.95 (0.17)	0.73 (0.03)
μ_{TOTAL}	1.90 (0.10)	6.70 (0.41)	2.59 (0.08)	2.17 (0.05)	6.43 (0.20)	2.78 (0.06)	2.25 (0.06)	6.47 (0.22)	2.85 (0.07)
U_{GENOME}	0.16	0.10	0.25	0.18	0.09	0.27	0.19	0.09	0.27

U_{GENOME} is the haploid genome-wide mutation rate per generation.

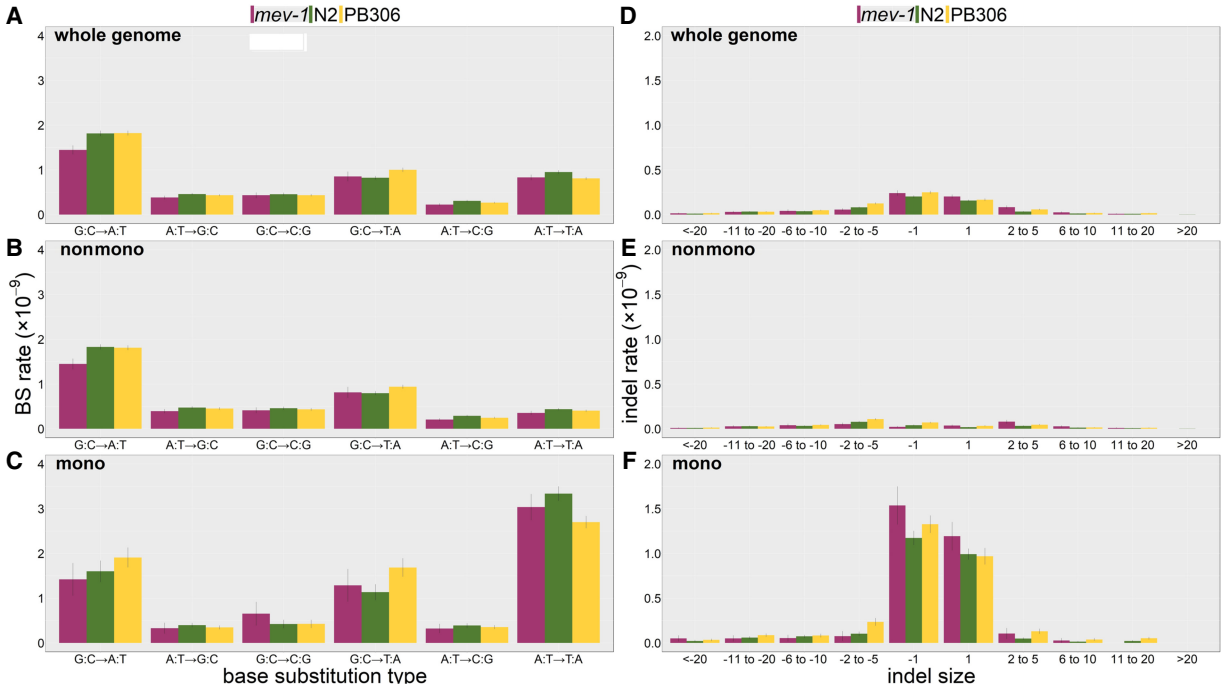

Figure 2. Mutation rates. (*A*–*C*) Type-specific base-substitution rates, by strain. (*D*–*F*) Size-specific indel rates, by strain. (*A*,*D*) Genome-wide mutation rates; (*B*,*E*) rates at nonmononucleotide repeat sequence; (*C*,*F*) mononucleotide repeat sequence. All rates are scaled as 10^{-9} per base per generation; note the difference in *y*-axis scale between panels *A*–*C* and panels *D*–*F*. Error bars, one SEM.

candidates that could potentially increase the mutation rate (see Supplemental Methods, section 4). It turns out there are obvious candidates in both lines (Supplemental Table S4). Line 572 has a T→G transversion upstream of the *atl-1* gene; *atl-1* is involved in homologous recombination and the DNA damage checkpoint. Line 481 has a 1-bp deletion downstream from the *xpc-1* gene; *xpc-1* is involved in nucleotide excision repair.

The mitochondrion is the site of the electron transport chain subunit II and therefore is the source location of the excess free radicals produced in *mev-1* worms (Senoo-Matsuda et al. 2001). However, the mtDNA mutation rate μ_{Mt} is not increased in *mev-1* and is very close to that of N2 (*mev-1*, $\mu_{Mt} = 5.62 \pm 2.62 \times 10^{-8}$/gen; N2, $\mu_{Mt} = 6.05 \pm 1.36 \times 10^{-8}$/gen; randomization test $P > 0.86$). The point estimate of the mtDNA mutation rate is ~50% greater in PB306 ($\mu_{Mt} = 8.84 \pm 2.15 \times 10^{-8}$/gen), but the difference between PB306 and N2 is not significantly different from zero (randomization test $P > 0.47$). A list of mtDNA mutations and their heteroplasmic frequencies is given in Supplemental Table S5. Details of the calculation of μ_{Mt} and of the randomization test are given in the Methods.

Mononucleotide repeats

Mononucleotide repeats are well known to incur indel mutations at a much greater rate than nonrepeat sequence (Denver et al. 2004a), and we previously found that mononucleotide repeats in the N2 genome of eight or more consecutive bases also experience an elevated rate of base-substitution mutations (Saxena et al. 2019). Here we find the same result when mononucleotides are defined by the less stringent criterion of five or more consecutive bases. Mononucleotide repeats of five or more bases comprise ~8.5% of the N2 genome, of which ~80% are A:T repeats (see Supplemental Methods, section 8). Pooled over strains and mutation types, the mononucleotide base-substitution rate is nearly twice that of nonmononucleotide sequence (Table 1). However, the increased base-substitution mutation rate in mononucleotide repeats is not uniform; the rate of transversions from either a G:C or an A:T base pair to an T:A base pair is greater in mononucleotide repeats, whereas rates of the other four types of base-substitution do not differ between sequence types (Fig. 2B,C). In particular, the rate of A:T→T:A transversions is approximately sevenfold greater in mononucleotide repeats than in nonrepeat sequence.

As expected, the rate of ±1-bp indel mutations is much greater in mononucleotide repeats than in nonrepeat sequence; averaged over strains, 1-bp deletions occur about 26 times more frequently and 1-bp insertions about 39 times more frequently in mononucleotides than in nonmononucleotide sequence. In contrast, the rate of both deletions and insertions >1 bp is only about twice as great in mononucleotides as in nonmononucleotide sequence (Fig. 2D–F; Table 1). Summed over sequence types and mutation types, the rate of 1-bp indels is ~25% greater in *mev-1* than in N2, whereas the rates in N2 and PB306 are similar.

Local sequence context

A previous analysis of the factors potentially affecting base-substitution mutability in N2 revealed a predominant role for the local (3-nucleotide [nt]) sequence context (Saxena et al. 2019). In that study, motifs with a C or G in the mutant (3′) site were more mutable on average than motifs with an A or T in the mutant site, with one exception: The triplet 5′-ttA-3′ was the most mutable of any of the 64 motifs, with the large majority of mutations being A:T→T:A transversions. Subsequent MA studies with the N2 strain of *C. elegans* have found a similarly high mutability of the 5′-tTa-3′ and 5′-tAa-3′ motifs, with the mutant site in the middle position

of a palindrome (Konrad et al. 2019; Volkova et al. 2020). The 64-nt triplets are depicted in Figure 3, oriented 5′ → 3′ with the mutant site in the middle position. There are not sufficiently many mutations to permit a robust statistical comparison between strains of the full set of 64 triplets, but we have a prior hypothesis with respect to the relative mutability of the 5′-tT/Aa-3′ triplets. 5′-tT/Aa-3′ are also the most mutable motifs in the other two strains. The 5′-tTAa-3′ and 5′-tAa-3′ mutation rates are similar to each other (Fig. 3) and differ among strains (N2, $\mu_{tT/Aa} = 8.3 \times 10^{-9}$/gen; *mev-1*, $\mu_{tT/Aa} = 7.0 \times 10^{-9}$/gen; PB306, $\mu_{tT/Aa} = 6.3 \times 10^{-9}$/gen); the difference between PB306 and N2 is statistically significant. The correlation of the 64 motif-specific mutation rates between strains is large and positive (>0.7) but is significantly less than one in all three cases (Supplemental Table S8).

Some fraction of 5′-tT/Aa-3′ triplets occur at the 3′ end of poly (T) repeats, suggesting the possibility that the atypically high mutability of those motifs results from the association of the motifs with mononucleotide repeats. Averaged over the three strains, $\mu_{tT/Aa}$ when the motif is adjacent to a poly(T) repeat is approximately 27-fold greater than when the same motif is not adjacent to a poly(T) repeat ($41.50 \pm 1.58 \times 10^{-9}$/generation vs. $1.55 \pm 0.12 \times 10^{-9}$/generation). When not in the context of a mononucleotide repeat, the 5′-*tT/Aa*-3′ motifs mutate at essentially the same rate as any A:T base pair (Fig. 3).

Details of the local sequence context analysis are presented in the Supplemental Methods, section 9.

Mutation spectrum

The mutation spectrum, defined as the frequency distribution of mutations of given types, is not the same as the distribution of type-specific mutation rates, although the two are obviously related. In our MA data, pooling over strains and the six base-substitution types, the correlation between the type-specific base-substitution mutation rate (μ_i) and the proportion of mutations of that type (p_i) is $r_{\mu,p} = 0.85$. The utility of knowing the mutation spectrum in MA lines is that it provides a benchmark for comparison to wild isolates, for which the mutation rate cannot be known.

Summed over all MA lines within strains, the six-category base-substitution mutation spectrum of *mev-1* ($n = 508$ mutations) does not differ from those of either N2 ($n = 3434$ mutations; 2×6 Monte Carlo Fisher's exact test, 10^6 iterations, $P > 0.50$) or PB306 ($n = 3111$ mutations; MC FET, $P > 0.45$), whereas the spectrum of N2 differs significantly from that of PB306 (MC FET, $P < 0.001$) (Fig. 4A). The differences between the base-substitution spectra of N2 and PB306 are not great, but the large number of mutations enables us to detect significant differences on the order of a few percent.

To compare the indel spectra between strains, we first investigated the overall mutational bias, quantified as the proportion of deletions among indel mutations. The indel bias differed significantly among strains (3×2 Fisher's exact test, $P < 0.04$). Breaking down the bias into pairwise comparisons between strains, *mev-1*

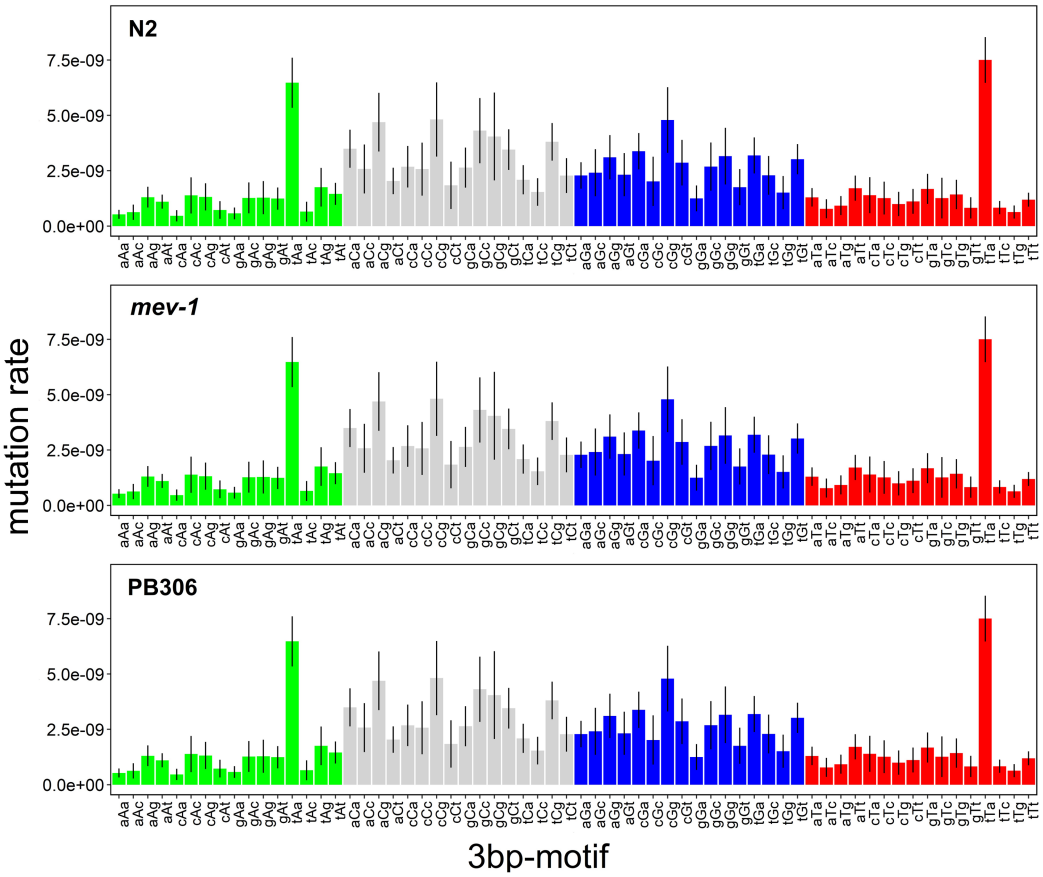

Figure 3. Sixty-four 3-bp motif base-substitution mutation rates. Motifs are arranged 5′-xYz-3′, with the mutant base Y in the middle position. Error bars, one SEM.

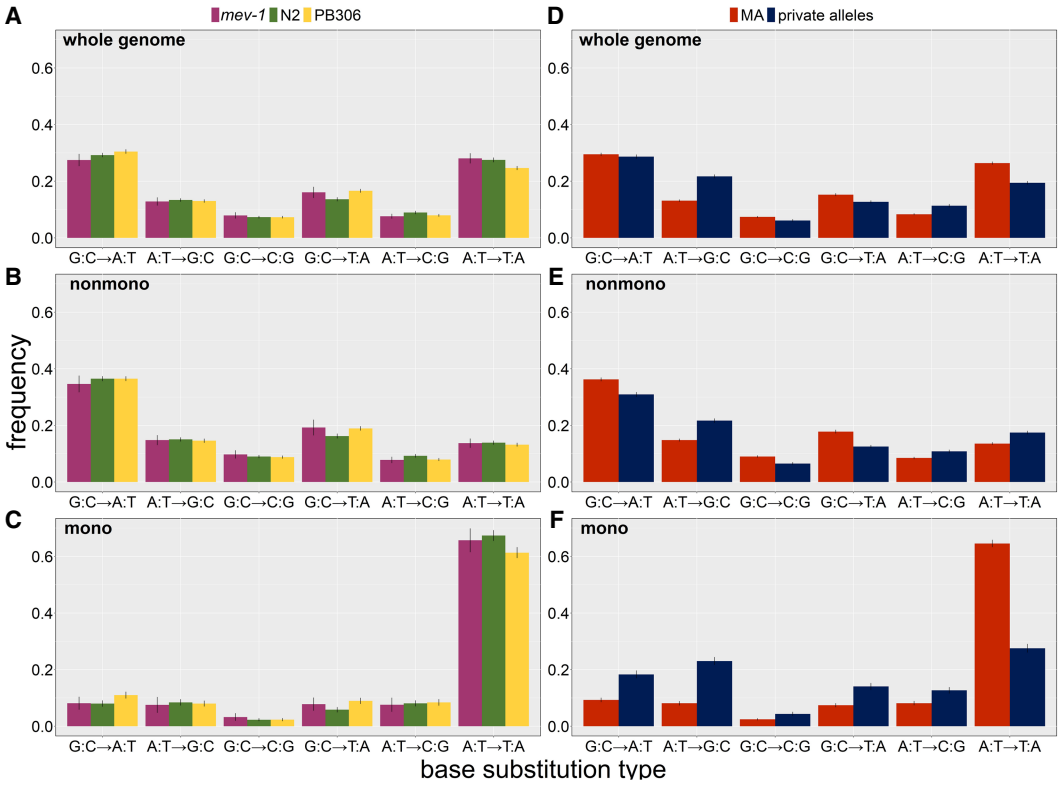

Figure 4. Base-substitution spectra. (*A*–*C*) MA lines; (*D*–*F*) wild isolate private alleles (blue) and MA means (red). (*A*,*D*) Whole genome; (*B*,*E*) nonmononucleotide sequence; (*C*,*F*) mononucleotide sequence. Error bars, one SEM.

(55% deletion) had a significantly lower proportion of deletions than N2 (63% deletion, Fisher's exact test, $P<0.04$) and PB306 (65% deletion, FET $P<0.02$); the bias did not differ significantly between N2 and PB306.

To characterize the indel spectra with finer resolution, we next assigned insertions and deletions separately to four bins of size one, two to five, six to 10, and 11 to 20. The bin sizes are obviously arbitrary, but the distribution provides cell counts of five or more observations for all but the largest insertion bin in the *mev-1* strain. The genome-wide spectrum (Fig. 5A) did not differ between *mev-1* and N2 (FET, $P>0.08$) or between N2 and PB306 (FET, $P>0.29$).

Breaking the genomic context into mononucleotide repeats and nonmononucleotides, the nonmononucleotide spectra (Fig. 5B) were marginally different between *mev-1* and N2 (FET, $P<0.03$) but not between N2 and PB306 (FET, $P>0.11$). The mononucleotide spectra (Fig. 5C) were very similar between *mev-1* and N2 (FET, $P>0.82$) but were significantly different between N2 and PB306 (FET, $P<0.01$). Averaged over N2 and PB306, the deletion bias is stronger in nonmononucleotide sequence (72.5% deletion) versus mononucleotides (58% deletion).

Comparison of the MA spectrum to the natural private allele spectrum

Base-substitutions

Variant alleles present in a single wild isolate ("private alleles") presumably arose as new mutations in the recent past and have been minimally scrutinized by selection. As expected from previous studies, the genome-wide base-substitution spectrum of unique variants is qualitatively different from those of the MA lines (Fig. 4D). Because of the large sample sizes, even small differences in proportions between groups are highly statistically significant (FET, $P<<0.0001$ in all cases), so we interpret the differences heuristically. We first consider the genome-wide spectrum and then parse the genome into mononucleotide repeats and nonmononucleotide repeats. The wild isolate spectra are compared with the average of the N2 and PB306 MA spectra (which are themselves significantly different). The full CeNDR wild isolate data set includes 773 unique genotypes, which collectively harbor more than 450,000 private single-nucleotide variants. Of those 773 wild isolates, we arbitrarily restricted our analysis to 444 isolates that carry 100 or fewer private alleles ($n=11{,}667$ private alleles), which represent about 400 generations of MA. A list of the wild isolates included and the set of private alleles for each isolate is given in Supplemental Table S9.

A:T → T:A transversions occur more frequently in the MA lines than in the wild isolates, and A:T → G:C transitions occur less frequently in the MA lines. G:C → T:A transversions, the original focus of this study, occur only slightly more frequently in the MA lines than among the rare variants. The wild isolate private allele spectrum is closer to that of PB306 (and *mev-1*) than it is to that of N2, but the MA spectra are clearly more similar to each other than either is to the wild isolate spectrum.

Comparison of the base-substitution spectra of mononucleotide repeat sequences to the spectra of nonmononucleotide sequences is revealing (Fig. 4B,C). For nonmononucleotide sequence (~90% of the genome), the private allele spectrum of wild isolates is similar to that of the MA lines, although the proportion of private alleles mutating from a C:G base pair is less than the

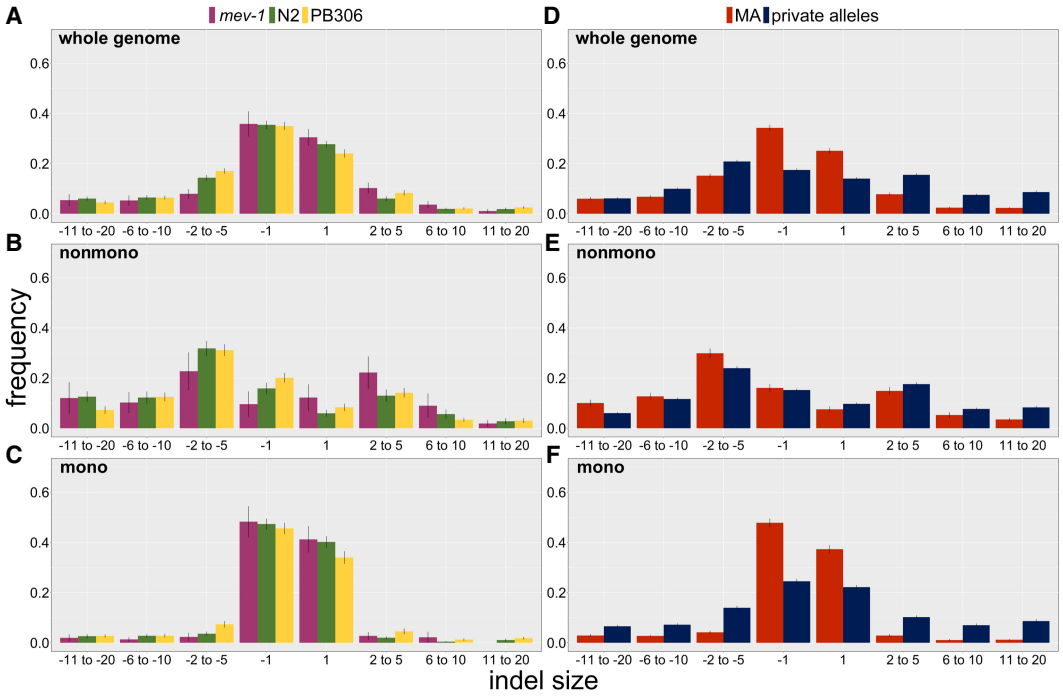

Figure 5. Indel spectra. (*A–C*) MA lines; (*D–F*) wild isolate private alleles (blue) and MA means (red). (*A,D*) Whole genome; (*B,E*) nonmononucleotide sequence; (*C,F*) mononucleotide sequence. Error bars, one SEM.

MA frequency in all three cases. In contrast, the mononucleotide spectra are very different. The proportion of A:T → T:A transversions in the wild isolates is less than half that expected from the MA proportion (28% vs. 65%), and there are many fewer A:T → G:C transitions in the MA lines than in the private alleles (8% vs. 25%).

If the strength of purifying selection varies among the different types of mutations, we expect the spectrum of common alleles will differ from that of rare variants. We compared the genome-wide allele frequency spectrum between private alleles and common(er) alleles, defined as variants present in more than one isolate but with an upper bound of 10% on the minor allele frequency. We chose 10% as the upper bound on the minor allele frequency to minimize the chance of misidentifying an ancestral allele as the recent mutant. The spectra of rare and common variants are similar (Supplemental Figs. S1, S2), reinforcing the notion that the difference in spectra between standing variants and mutations accumulated in laboratory MA experiments results from a consistent difference between the mutational milieu in the laboratory and in nature, rather than from the action of differential purifying selection.

As noted, the MA spectrum differs slightly but significantly between N2 and PB306. N2 is long-adapted to the laboratory environment and differs from wild-type *C. elegans* in many respects (Sterken et al. 2015), of which the mutation spectrum is but one. Private alleles in the wild isolates provide a much broader sample of the mutational process in the species. Parametric bootstrap simulations (for details, see Methods) show that there is more variation in the spectrum of private alleles among wild isolates than expected from a uniform base-substitution mutational process (Supplemental Fig. S3), for both nonmononucleotide sequence and mononucleotide runs, as well as genome-wide. This finding shows that the *C. elegans* mutational process varies among strains in nature as well as in the laboratory, although the causes of the variation in nature are uncertain.

Indels

The frequencies of false positives (FPs) and false negatives (FNs) are qualitatively greater for indels of >20 bp than for smaller indels (see next section), so we restrict comparison of indel spectra of MA lines and wild isolates to indels of <21 bp. To compare the indel spectra of MA lines and natural private alleles, we applied additional filters on the MA lines and used the full set of 773 wild isolates (Supplemental Methods, section 11).

As noted previously, the MA lines have a strong genome-wide deletion bias (~62% deletion); the deletion bias in the wild isolates is much weaker (54% deletion). Inspection of the genome-wide spectrum (Fig. 5D) reveals that ±1-bp indels are a much greater fraction of indels in the MA lines than in the wild isolates. Breaking the genome into mononucleotide and nonmononucleotide components, it is apparent that the nonmononucleotide spectra are roughly congruent between MA lines and wild isolates (Fig. 5E). In contrast, the mononucleotide spectra (Fig. 5F) are different between MA and wild isolates, with the MA lines having a greater fraction of ±1-bp indels. Another study with a different set of N2 strain MA lines reported a similarly high frequency of ±1-bp indels in mononucleotide runs (Konrad et al. 2019, see their Figure 8). Evidently, the genome-wide discrepancy between the MA indel spectrum and the private allele spectrum of wild isolates is largely owing to the different mutational properties of mononucleotide repeats in laboratory populations and in nature.

FPs and FNs

For an estimate of mutation rate to be believable, it is necessary to have credible estimates of the frequency of FPs (apparent variants

called as new mutations that are not new mutations) and FNs (mutated sites in the genome that are not called as new mutations). We designate alleles that are homozygous in the *C. elegans* reference genome (N2) as "0"; variant alleles are designated as "1." Some sites in the common ancestors of our MA lines differ from the reference allele; that is, the MA ancestor genotype is 1/1 (1439 sites in the N2 ancestor, 1642 in the *mev-1* ancestor, 155,614 in the PB306 ancestor).

FPs were assessed in two ways. First, we counted sites that were scored 1/1 in our 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 (alternatively, they could be errors in the reference genome). Then, we scored those sites in each MA line in the set. If a site was scored 1/1 in the MA ancestor but scored 0/0 in all MA lines, we inferred a FP in the MA ancestor. If the number of FPs resulting from sequencing errors depends only on the number of bases sequenced, the number should be similar for N2 and PB306. The number of true positives in this case is the number of sites at which the MA ancestor differs from the reference genome. The FDR for N2 is FP/(FP + 1439); the FDR for PB306 is FP/(FP + 155,614).

In N2, of the 1439 sites that differ between our G0 ancestor and the reference genome, we found one FP, a G:C → T:A transversion in a mononucleotide region. Of the 155,614 sites that differ between PB306 and the (N2) reference genome, we found four FPs for base-substitutions, two of which were in mononucleotide repeats (a G:C → A:T transition and a A:T → T:A transversion) and two were A:T → T:A transversions in nonmononucleotide sequence. The genome-wide false-discovery rate (FDR = FP/[FP + TP]) in the N2 MA lines for base-substitution mutations is 0.19%, and in PB306, it is 0.0033%. The mononucleotide base-substitution FDRs are 1.14% (N2) and 0.0096% (PB306). Genome-wide FDRs for indels in the two strains are zero and 0.28%, respectively. In PB306, 5/10 FPs were in mononucleotide repeats (0.023%) and 5/10 were in nonmononucleotides (0.035%).

Second, we used 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 in Supplemental Methods, section 12; RIAIL genotypes are given in Supplemental Table S10. Each parental line carries its own set of putative mutations, which are expected to segregate in the RIAILs with an average frequency of 50%. A variant called as a mutant in a parental line (1/1) that does not segregate in the RIAILs is designated as a FP in the parent. The FDR for detection of SNPs and indels is 2.1% and 2.9%, respectively. This result is consistent with a previous estimate of the FDR from N2 strain MA lines based on Sanger sequencing of PCR products, which gave an upper 95% confidence limit on the FDR of 2.5% (Saxena et al. 2019).

True mutations may not be identified as such for two reasons. First, the mutant site may not be covered by the sequencing. We refer to this situation as "failure to recall" a mutation. Second, a true mutation (i.e., 0/0 in the ancestor and 1/1 in the MA line carrying the mutation) may be incorrectly called as "not 1/1," that is, either as 0/0 or as a heterozygote, 0/1. These are FNs. FNs will cause the mutation rate to be underestimated. Failure to recall mutations will only cause the mutation rate to be misestimated (under or over) if the mutation rate at sites not covered in the sequencing differs from that of the sites covered.

We used a simulation approach to assess recall rate, the details of which are given in the Supplemental Methods, section 10. "Dummy" mutations were introduced into the reference genome at random (Supplemental Table S12; Supplemental Fig. S12), and the MA data were analyzed using our standard variant-calling pipeline but with the simulated genome as the reference. Sites with dummy mutations are scored as 1/1 in the MA ancestor and all MA lines. If a dummy mutation at a site was not called 1/1 in an MA line for any reason, it was classified as a failure to recall. The genome-wide base-substitution failure-to-recall rate is 6.81% in N2, 6.92% in *mev-1*, and 9.18% in PB306. The small indel (≤50-bp) failure-to-recall rates are 9.34% in N2, 9.78% in *mev-1*, and 12.35% in PB306.

Dummy mutations that are called as "not 1/1" (i.e., 0/0 or 0/1) are FNs *sensu stricto*. The genome-wide base-substitution FN rate (FNR) is 0.12% in N2, 0.17% in *mev-1*, and 0.91% in PB306. The FNR for mononucleotide base-substitutions in the N2 MA lines is about two times that of nonmononucleotide repeats but is still low (0.21% vs. 0.11%). The overall FNRs in PB306 are slightly higher and do not differ between mononucleotides (0.95%) and nonmononucleotides (0.91%). The small indel (≤50-bp) FNRs are 5.15% in N2, 5.59% in *mev-1*, and 5.84% in PB306.

Discussion

In this study, four key findings emerge. First, the base-substitution mutational process of *mev-1* is similar to that of N2, from which it was derived. Therefore, any discrepancy between the base-substitution spectrum of N2 and those of wild isolates is unlikely to be owing to differences resulting from endogenous oxidative stress manifested in the laboratory environment. Moreover, the overall base-substitution rate of *mev-1* is significantly less (~14%) than that of N2, contrary to the prediction that elevated levels of ROS increase the base-substitution mutation rate. In contrast, the indel rate, and especially the insertion rate, is greater in the *mev-1* lines than in N2 (~50% greater). The dinucleotide repeat indel rate increases in N2 MA lines propagated under conditions of heat stress (Matsuba et al. 2013), and a similar increase in the indel rate—but not the base-substitution rate—has been shown in *Arabidopsis thaliana* MA lines cultured under conditions of heat stress (Belfield et al. 2021). *Drosophila melanogaster* MA lines initiated from genomes carrying deleterious large-effect mutations experience higher rates of short deletions than wild-type MA lines, thought to be owing to preferential deployment of different mechanisms of double-strand break repair (Sharp and Agrawal 2016). The increased insertion rate in the *mev-1* lines suggests that elevated levels of endogenous oxidative damage may in fact influence the mutational process, but not in the way that we predicted.

A possible explanation for the lower base-substitution rate in *mev-1* relative to N2 is that the N2 lines underwent approximately twice as many generations of MA (Gmax = 250 generations) as did *mev-1* (Gmax = 125 generations). We previously found that "second-order" N2 MA lines initiated from a subset of the N2 lines included in this study evolved a significantly greater (~10%) mutation rate over a subsequent approximately 150 generations of MA. The results reported here are consistent with the N2 (and PB306) lines evolving an increased base-substitution mutation rate over time, although the evidence is circumstantial.

The second key result emerges from the comparison of the mutational properties of the laboratory-adapted N2 strain and the wild isolate PB306. The overall base-substitution mutation rates are very similar between the two strains (<5% different),

although some of the type-specific mutation rates are marginally different (e.g., the genome-wide G:C → A:T transition rate is ~20% greater in PB306, and the A:T → T:A transversion rate is ~20% greater in N2). The type-specific differences between strains are not large; by way of contrast, two sets of *D. melanogaster* MA lines derived from different starting genotypes had a fivefold difference in G:C → A:T transition rates (Schrider et al. 2013). PB306 has a similarly greater indel rate than N2 (~25%), although the direction and magnitude of the bias are nearly identical (65% and 63% deletion, respectively).

The differences in mutational properties between N2 and PB306 surely have a genetic basis. The two strains' genomes are ~1.5% different (approximately 150,000 pairwise differences). With only two strains, we cannot say whether the difference is related to the domestication of N2, although it is interesting that the mutational properties of *mev-1* trend in the same direction as those of PB306. If the elevated endogenous oxidative stress experienced by *mev-1* is reflective of environmental stress in general, it suggests that the laboratory environment may be slightly stressful for PB306 compared with the laboratory-adapted N2. Moreover, the private allele spectrum of the wild isolates is closer to that of PB306 than of N2, which, by the preceding logic, suggests that the laboratory environment may be especially comfortable for N2 relative to the natural environment. The finding that there is too much variation in the standing mutation spectrum among wild isolates to be explained by a uniform mutational process suggests that the genetic variation in the mutational process observed between N2 and PB306 is not an artifact of the laboratory environment. Probably the most important point to take away from the comparison, however, is that both the rate and spectra of N2 and PB306 are not very different, which implies that conclusions about mutation drawn from N2 MA lines may be generalizable to *C. elegans* at large.

The third key finding is the discrepancy in the indel bias between the MA lines and the wild isolate private alleles: MA lines show a strong deletion bias (~62% genome-wide), whereas the deletion bias in the private alleles is much weaker (~54% deletion). As noted above, the indel rate has been shown to increase under stress in a variety of experimental systems, including *C. elegans*. The reduced deletion bias in *mev-1* relative to N2 reinforces the inference that a nematode's life in nature is more stressful than life in the laboratory, with the attendant differences in the mutational process. To the extent that mutational variance in phenotypic traits is underlain by indel mutations, evolutionary inferences that rely on comparisons between mutational and standing genetic variance are rendered less robust.

The fourth key finding is the qualitative difference in the base-substitution mutation spectra between mononucleotide repeat sequence and nonmononucleotide sequence in the MA lines, as well as how the MA spectra relate to the standing private allele spectra. This study was predicated on the discrepancy between the (N2) MA base-substitution spectrum and that seen in wild isolates. In fact, when the nonmononucleotide spectra are compared, the discrepancy substantially disappears. Although significant differences remain, they do not appear wildly different (Fig. 4E). Conversely, there is a huge discrepancy between the mononucleotide spectra of MA lines and private alleles and, in particular, with respect to how A:T mutates (Fig. 4F). First, a larger fraction of mutations occur at A:T sites in the MA lines than in the private alleles (81% vs. 66%). Second, there are many more A:T → T:A transversions in the MA lines than in the private alleles (65% vs. 28%), and there are many fewer A:T → G:C transitions in the MA lines than in the private alleles (8% vs. 25%). Variants at mononucleotide loci are more prone to both FPs and (especially) FNs. Base-substitution FNRs in the MA lines are low (<1% in mononucleotides). We cannot discriminate "true" FPs in MA lines or RIAILs from FNs in the ancestor genome, but if we conservatively assume that all inferred FPs are "true" FPs, the FPR is still low (≤2%), even in mononucleotide repeats. The technology used in sequencing most (not all) of the wild isolates and most (not all) of the MA lines was the same, and the data were analyzed in the same way (Supplemental Methods, sections 4 and 13; Supplemental Figs. S4–S11; Supplemental Table S11), so it is difficult to imagine that the discrepancy is owing to wildly different rates of FPs or FNs in one or the other set of data. Given that the genetic basis of eukaryotic DNA repair is complex and that hundreds of genes are known to influence variation in DNA repair (Eisen and Hanawalt 1999), a reasonable guess is that some element(s) of DNA repair differs systematically between the laboratory environment and the natural environment, but additional studies are needed to determine what that may be.

Methods

Strains

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* allele 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, 2011, 2013, 2016) and includes elevated 8-oxodG and increased nuclear mutation rate (Hartman et al. 2004; Ishii et al. 2005). The *mev-1* allele was backcrossed into the canonical Baer laboratory N2 strain genomic background (for details, see Supplemental Methods).

PB306 is a wild isolate generously provided by Scott Baird; the N2 lines are derived from a replicate of the ancestor of the Vassilieva and Lynch (1999) MA lines.

Wild isolates

Genome sequence data for 773 unique wild isolates were obtained from the *C. elegans* Natural Diversity Resource (CeNDR; Cook et al. 2017). We chose 444 isolates for further analysis, based on the criterion that an isolate carry 100 or fewer private alleles. The list of wild isolates included in the analysis is given in Supplemental Table S9. Details of the sequencing and variant calling are given in sections 3 and 4 of the Supplemental Methods. Collection data are available at https://www.elegansvariation.org/.

MA experiments

Details of the N2 and PB306 MA experiment are given by Baer et al. (2005); those of the *mev-1* MA experiment are given by Joyner-Matos et al. (2011) and are summarized in the Supplemental Methods, section 2. Data on MA line transfers are presented in Supplemental Table S1. *mev-1* MA lines were propagated for up to 125 generations of MA (Gmax = 125); N2 and PB306 lines were propagated for up to 250 generations of MA (Gmax = 250). The basic protocol in both experiments follows that of Vassilieva and Lynch (1999) and is depicted in Figure 1.

Genome sequencing and bioinformatics

Thirty-two N2 lines and 20 PB306 lines were previously sequenced with Illumina short-read sequencing at an average coverage of ~ 27× read-depth with 100-bp paired-end reads. The N2 data have been previously reported (Saxena et al. 2019). For this project, we sequenced an additional 40 N2 lines, 54 PB306 lines, and 24 *mev-1* lines and their ancestors using Illumina short-read sequencing at an average coverage of 46× depth with 150-bp paired-end reads. Protocols for DNA extraction and construction of sequencing libraries are given in the Supplemental Methods, section 3; details of bioinformatics processing of raw sequence data and variant calling are given in the Supplemental Methods, section 4.

Following preliminary analysis, variants were called using HaplotypeCaller (in BP_RESOLUTION mode) in GATK4 (v4.1.4.0) (McKenna et al. 2010). Variants were identified as putative mutations if (1) the variant was identified as homozygous and (2) it was present in one and only one MA line. Criterion 1 means that any mutations that occurred in the last few generations of MA that were still segregating and/or occurred during population expansion for DNA extraction were ignored. Because the MA progenitor was at mutation-drift equilibrium (Lynch and Hill 1986), the segregating variation is expected to be the same in the MA progenitor and the MA line, so ignoring heterozygotes results in an unbiased estimate of mutation rate. Criterion 2 reduces the probability of mistakenly identifying a variant segregating at low frequency in the expanded population of the MA progenitor as a new mutation. One *mev-1* line was found not to carry the *mev-1* allele and was omitted as a presumed contaminant. Four pairs of N2 strain MA lines and seven pairs of PB306 strain lines shared multiple variants and were inferred to have experienced contamination at some point during the MA phase; one line from each pair was arbitrarily omitted from subsequent analyses. The final data set includes 68 N2 lines, 67 PB306 lines, and 23 *mev-1* lines (Supplemental Table S3).

Data analysis

Nuclear mutation rate

Because different lines experienced different numbers of generations during the MA phase and because the number of callable sites differs among lines, we cannot use the number of mutations per line to directly compare mutation rates among strains. Instead, we calculated a mutation rate, μ_i, for each 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). To test for group-specific effects on mutation rate and to partition the (co)variance in mutation rate into within- and among-group components, we consider the mutation rate itself as a continuously distributed dependent variable in a GLM. Details of the GLM analyses are given in the Supplemental Methods, section 5.

mtDNA mutation rate

Estimation of the mtDNA mutation rate is more complicated than for nuclear loci because a nontrivial 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 jth 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). However, because many lines have no mtDNA mutations, GLM is not an appropriate analysis. Instead, we used randomization tests of hypothesized differences in mtDNA mutation rate; the details are presented in the Supplemental Methods, section 6.

Mutation spectrum

Mutation spectra were compared between groups by Fisher's exact test. For comparisons with sample sizes too large to calculate directly we used Monte Carlo estimates of the FET as implemented in the FREQ procedure of SAS v.9.4.

We also tested 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 used parametric bootstrap simulations, the details of which are presented in the Supplemental Methods, section 7.

Data access

All raw sequencing data in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession numbers PRJNA429972 (23 N2 MA lines), PRJNA395568 (nine N2 MA lines), and PRJNA665851 (all other MA lines). All code for variant calling is available in the Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank J. Dembek for laboratory assistance and the anonymous reviewers for their helpful suggestions. Support was provided by National Institutes of Health awards GM107227 to C.F.B. and E.C.A. and GM127433 to C.F.B. and V. Katju.

References

Babbitt GA, Cotter CR. 2011. Functional conservation of nucleosome formation selectively biases presumably neutral molecular variation in yeast genomes. *Genome Biol Evol* **3:** 15–22. doi:10.1093/gbe/evq081

Baer CF. 2019. Evolution: environmental dependence of the mutational process. *Curr Biol* **29:** R415–R417. doi:10.1016/j.cub.2019.04.049

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. *Proc Natl Acad Sci USA* **102:** 5785–5790. doi:10.1073/pnas.0406056102

Beckman KB, Ames BN. 1998. The free radical theory of aging matures. *Physiol Rev* **78:** 547–581. doi:10.1152/physrev.1998.78.2.547

Belfield EJ, Brown C, Ding ZJ, Chapman L, Luo M, Hinde E, van Es SW, Johnson S, Ning Y, Zheng SJ, et al. 2021. Thermal stress accelerates *Arabidopsis thaliana* mutation rate. *Genome Res* **31:** 40–50. doi:10.1101/gr.259853.119

Bridge G, Rashid S, Martin SA. 2014. DNA mismatch repair and oxidative DNA damage: implications for cancer biology and treatment. *Cancers (Basel)* **6:** 1597–1614. doi:10.3390/cancers6031597

Busuttil RA, Rubio M, Dollé MET, Campisi J, Vijg J. 2003. Oxygen accelerates the accumulation of mutations during the senescence and immortalization of murine cells in culture. *Aging Cell* **2:** 287–294. doi:10.1046/j.1474-9728.2003.00066.x

Cadet J, Douki T, Gasparutto D, Ravanat J-L. 2003. Oxidative damage to DNA: formation, measurement and biochemical features. *Mut Res* **531:** 5–23. doi:10.1016/j.mrfmmm.2003.09.001

Carlson J, Locke AE, Flickinger M, Zawistowski M, Levy S, Myers RM, Boehnke M, Kang HM, Scott LJ, Li JZ, et al. 2018. Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *Nat Commun* **9:** 3753. doi:10.1038/s41467-018-05936-5

Cheng KC, Cahill DS, Kasai H, Nishimura S, Loeb LA. 1992. 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes

G-T and A-C substitutions. *J Biol Chem* **267:** 166–172. doi:10.1016/S0021-9258(18)48474-8
Constantini D. 2014. *Oxidative stress and hormesis in evolutionary ecology and physiology*. Springer, New York.
Cook DE, Zdraljevic S, Roberts JP, Andersen EC. 2017. CeNDR, the *Caenorhabditis elegans* Natural Diversity Resource. *Nucleic Acids Res* **45:** D650–D657. doi:10.1093/nar/gkw893
Cooke MS, Evans MD, Dizdaroglu M, Lunec J. 2003. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J* **17:** 1195–1214. doi:10.1096/fj.02-0752rev
Cunningham RP. 1997. DNA repair: caretakers of the genome? *Curr Biol* **7:** R576–R579. doi:10.1016/S0960-9822(06)00286-7
Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289:** 2342–2344. doi:10.1126/science.289.5488.2342
Denver DR, Morris K, Kewalramani A, Harris KE, Chow A, Estes S, Lynch M, Thomas WK. 2004a. Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. *J Mol Evol* **58:** 584–595. doi:10.1007/s00239-004-2580-4
Denver DR, Morris K, Lynch M, Thomas WK. 2004b. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430:** 679–682. doi:10.1038/nature02697
Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc Natl Acad Sci USA* **106:** 16310–16314. doi:10.1073/pnas.0904895106
Denver DR, Wilhelm LJ, Howe DK, Gafner K, Dolan PC, Baer CF. 2012. Variation in base-substitution mutation in experimental and natural lineages of *Caenorhabditis* nematodes. *Genome Biol Evol* **4:** 513–522. doi:10.1093/gbe/evs028
Dollé MET, Snyder WK, Gossen JA, Lohman PHM, Vijg J. 2000. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc Natl Acad Sci USA* **97:** 8403–8408. doi:10.1073/pnas.97.15.8403
Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. *Genetics* **148:** 1667–1686. doi:10.1093/genetics/148.4.1667
Eisen JA, Hanawalt PC. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* **435:** 171–213. doi:10.1016/S0921-8777(99)00050-6
Evans MD, Cooke MS. 2004. Factors contributing to the outcome of oxidative damage to nucleic acids. *Bioessays* **26:** 533–542. doi:10.1002/bies.20027
Gao ZY, Moorjani P, Sasani TA, Pedersen BS, Quinlan AR, Jorde LB, Amster G, Przeworski M. 2019. Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc Natl Acad Sci USA* **116:** 9491–9500. doi:10.1073/pnas.1901259116
Gray JM, Karow DS, Lu H, Chang AJ, Chang JS, Ellis RE, Marletta MA, Bargmann CI. 2004. Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430:** 317–322. doi:10.1038/nature02714
Guo C, McDowell IC, Nodzenski M, Scholtens DM, Allen AS, Lowe WL, Reddy TE. 2017. Transversions have larger regulatory effects than transitions. *BMC Genomics* **18:** 394. doi:10.1186/s12864-017-3785-4
Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, Gunnarsson B, Oddsson A, Halldorsson GH, Zink F, et al. 2019. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363:** eaau1043. doi:10.1126/science.aau1043
Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Ann Rev Ecol Evol Syst* **40:** 151–172. doi:10.1146/annurev.ecolsys.39.110707.173437
Halliwell B, Gutteridge JMC. 2007. *Free radicals in biology and medicine*. Oxford University Press, London.
Hartman P, Ponder R, Lo HH, Ishii N. 2004. Mitochondrial oxidative stress can lead to nuclear hypermutability. *Mech Ageing Dev* **125:** 417–420. doi:10.1016/j.mad.2004.02.007
Helbock HJ, Beckman KB, Shigenaga MK, Walter PB, Woodall AA, Yeo HC, Ames BN. 1998. DNA oxidation matters: the HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine. *Proc Natl Acad Sci USA* **95:** 288–293. doi:10.1073/pnas.95.1.288
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*. *Mutat Res* **237:** 165–171. doi:10.1016/0921-8734(90)90022-J
Ishii N, Fujii M, Hartman PS, Tsuda M, Yasuda K, Senoo-Matsuda N, Yanase S, Ayusawa D, Suzuki K. 1998. A mutation in succinate dehydrogenase cytochrome *b* causes oxidative stress and ageing in nematodes. *Nature* **394:** 694–697. doi:10.1038/29331
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, Miyazawa M, Hartman PS, Ishii N. 2011. Mitochondrial superoxide anion (O_2^-) inducible "*mev-1*" animal models for aging research. *BMB Rep* **44:** 298–305. doi:10.5483/BMBRep.2011.44.5.298
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. doi:10.1016/j.bbabio.2012.10.016
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. doi:10.1016/j.mad.2016.02.013
Joyner-Matos J, Bean LC, Richardson HL, Sammeli T, Baer CF. 2011. No evidence of elevated germline mutation accumulation under oxidative stress in *Caenorhabditis elegans*. *Genetics* **189:** 1439–1447. doi:10.1534/genetics.111.133660
Katju V, Bergthorsson U. 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 Biol Evol* **11:** 136–165. doi:10.1093/gbe/evy252
Keightley PD, Caballero A. 1997. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* **94:** 3823–3827. doi:10.1073/pnas.94.8.3823
Kessler MD, Loesch DP, Perry JA, Heard-Costa NL, Taliun D, Cade BE, Wang H, Daya M, Ziniti J, Datta S, et al. 2020. De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the amish founder population. *Proc Natl Acad Sci USA* **117:** 2560–2569. doi:10.1073/pnas.1902766117
Kim S-Y, Velando A. 2020. Attractive male sticklebacks carry more oxidative DNA damage in the soma and germline. *J Evol Biol* **33:** 121–126. doi:10.1111/jeb.13552
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. *Mol Biol Evol* **34:** 1319–1334. doi:10.1093/molbev/msx051
Konrad A, Brady MJ, Bergthorsson U, Katju V. 2019. Mutational landscape of spontaneous base substitutions and small indels in experimental *Caenorhabditis elegans* populations of differing size. *Genetics* **212:** 837–854. doi:10.1534/genetics.119.302054
Kašovec R, Richards H, Gifford DR, Hatcher C, Faulkner KJ, Belavkin RV, Channon A, Aston E, McBain AJ, Knight CG. 2017. Spontaneous mutation rate is a plastic trait associated with population density across domains of life. *PLoS Biol* **15:** e2002731. doi:10.1371/journal.pbio.2002731
Kreuz S, Fischle W. 2016. Oxidative stress signaling to chromatin in health and disease. *Epigenomics* **8:** 843–862. doi:10.2217/epi-2016-0002
Liu HX, Zhang JZ. 2019. Yeast spontaneous mutation rate and spectrum vary with environment. *Curr Biol* **29:** 1584–1591.e3. doi:10.1016/j.cub.2019.03.054
Lynch M, Hill WG. 1986. Phenotypic evolution by neutral mutation. *Evolution (NY)* **40:** 915–935. doi:10.1111/j.1558-5646.1986.tb00561.x
Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci USA* **90:** 4087–4091. doi:10.1073/pnas.90.9.4087
Matsuba C, Ostrow DG, Salomon MP, Tolani A, Baer CF. 2013. Temperature, stress and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Biol Lett* **9:** 20120334. doi:10.1098/rsbl.2012.0334
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 Res* **20:** 1297–1303. doi:10.1101/gr.107524.110
Messer PW. 2009. Measuring the rates of spontaneous mutation from deep and large-scale polymorphism data. *Genetics* **182:** 1219–1232. doi:10.1534/genetics.109.105692
Mukai T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I: spontaneous mutation rate of polygenes controlling viability. *Genetics* **50:** 1–19. doi:10.1093/genetics/50.1.1
Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76:** 5269–5273. doi:10.1073/pnas.76.10.5269
Paul C, Nagano M, Robaire B. 2011. Aging results in differential regulation of DNA repair pathways in pachytene spermatocytes in the Brown Norway rat. *Biol Repro* **85:** 1269–1278. doi:10.1095/biolreprod.111.094219
Poetsch AR, Boulton SJ, Luscombe NM. 2018. Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome Biol* **19:** 215. doi:10.1186/s13059-018-1582-2

Richter C, Park JW, Ames BN. 1988. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci USA* **85:** 6465–6467. doi:10.1073/pnas.85.17.6465

Saxena AS, Salomon MP, Matsuba C, Yeh SD, Baer CF. 2019. Evolution of the mutational process under relaxed selection in *Caenorhabditis elegans*. *Mol Biol Evol* **36:** 239–251. doi:10.1093/molbev/msy213

Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* **194:** 937–954. doi:10.1534/genetics.113.151670

Senoo-Matsuda N, Yasuda K, Tsuda M, Ohkubo T, Yoshimura S, Nakazawa H, Hartman PS, Ishii N. 2001. A defect in the cytochrome *b* large subunit in complex II causes both superoxide anion overproduction and abnormal energy metabolism in *Caenorhabditis elegans*. *J Biol Chem* **276:** 41553–41558. doi:10.1074/jbc.M104718200

Sharp NP, Agrawal AF. 2012. Evidence for elevated mutation rates in low-quality genotypes. *Proc Natl Acad Sci USA* **109:** 6142–6146. doi:10.1073/pnas.1118918109

Sharp NP, Agrawal AF. 2016. Low genetic quality alters key dimensions of the mutational spectrum. *PLoS Biol* **14:** e1002419. doi:10.1371/journal.pbio.1002419

Shigenaga MK, Gimeno CJ, Ames BN. 1989. Urinary 8-hydroxy-2′-deoxyguanosine as a biological marker of in vivo oxidative DNA damage. *Proc Natl Acad Sci USA* **86:** 9697–9701. doi:10.1073/pnas.86.24.9697

Sprague GF, Russell WA, Penny LH. 1960. Mutations affecting quantitative traits in the selfed progeny of doubled monoploid maize stocks. *Genetics* **45:** 855–866. doi:10.1093/genetics/45.7.855

Sterken MG, Snoek LB, Kammenga JE, Andersen EC. 2015. The laboratory domestication of *Caenorhabditis elegans*. *Trends Genet* **31:** 224–231. doi:10.1016/j.tig.2015.02.009

Stoltzfus A. 2008. Evidence for a predominant role of oxidative damage in germline mutation in mammals. *Mut Res* **644:** 71–73. doi:10.1016/j.mrfmmm.2008.05.003

Suzuki T, Kamiya H. 2017. Mutations induced by 8-hydroxyguanine (8-oxo-7,8-dihydroguanine), a representative oxidized base, in mammalian cells. *Genes Environ* **39:** 2. doi:10.1186/s41021-016-0051-y

Turrens JF. 2003. Mitochondrial formation of reactive oxygen species. *J Physiol* **552:** 335–344. doi:10.1113/jphysiol.2003.049478

Vassilieva LL, Lynch M. 1999. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151:** 119–129. doi:10.1093/genetics/151.1.119

Volkova NV, Meier B, Gonzalez-Huici V, Bertolini S, Gonzalez S, Vohringer H, Abascal F, Martincorena I, Campbell PJ, Gartner A, et al. 2020. Mutational signatures are jointly shaped by DNA damage and repair. *Nat Commun* **11:** 2169. doi:10.1038/s41467-020-15912-7

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7:** 256–276. doi:10.1016/0040-5809(75)90020-9

Wright S. 1931. Evolution in Mendelian populations. *Genetics* **16:** 97–159. doi:10.1093/genetics/16.2.97

Received February 9, 2021; accepted in revised form July 12, 2021.