Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in *Caenorhabditis elegans* Spontaneous Mutation Accumulation Lines of Differing Population Size

Anke Konrad,[1] Owen Thompson,[2] Robert H. Waterston,[2] Donald G. Moerman,[3] Peter D. Keightley,[4] Ulfar Bergthorsson,*[,†,1] and Vaishali Katju*[,†,1]

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX

[2]Department of Genome Sciences, University of Washington, Seattle, WA

[3]Department of Zoology and Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada

[4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: ubergthorsson@cvm.tamu.edu; vkatju@cvm.tamu.edu.

Associate editor: Jeffrey P. Townsend

Abstract

Mitochondrial genomes of metazoans, given their elevated rates of evolution, have served as pivotal markers for phylogeographic studies and recent phylogenetic events. In order to determine the dynamics of spontaneous mitochondrial mutations in small populations in the absence and presence of selection, we evolved mutation accumulation (MA) lines of *Caenorhabditis elegans* in parallel over 409 consecutive generations at three varying population sizes of $N = 1$, 10, and 100 hermaphrodites. The $N = 1$ populations should have a minimal influence of natural selection to provide the spontaneous mutation rate and the expected rate of neutral evolution, whereas larger population sizes should experience increasing intensity of selection. New mutations were identified by Illumina paired-end sequencing of 86 mtDNA genomes across 35 experimental lines and compared with published genomes of natural isolates. The spontaneous mitochondrial mutation rate was estimated at 1.05×10^{-7}/site/generation. A strong G/C→A/T mutational bias was observed in both the MA lines and the natural isolates. This suggests that the low G + C content at synonymous sites is the product of mutation bias rather than selection as previously proposed. The mitochondrial effective population size per worm generation was estimated to be 62. Although it was previously concluded that heteroplasmy was rare in *C. elegans*, the vast majority of mutations in this study were heteroplasmic despite an experimental regime exceeding 400 generations. The frequencies of frameshift and nonsynonymous mutations were negatively correlated with population size, which suggests their deleterious effects on fitness and a potent role for selection in their eradication.

***Key words:* *Caenorhabditis elegans*, mitochondrial effective population size, genetic drift, heteroplasmy, mitochondria, mutation rate, selection, spontaneous mutation.**

Article

Introduction

The organellar genes and genomes of mitochondria and chloroplasts have been the subject of intensive evolutionary investigations over the past three decades with respect to genetic variation and population-genetic analyses. The transmission genetics of these organelles is inherently different from nuclear genes with respect to (i) their mutation rates, (ii) their patterns of segregation, (iii) their mode of inheritance, and (iv) the degree of intracellular selection and recombination (Birky 2001). For example, animal mtDNA possesses a high mutation rate, approximately a magnitude greater than that of nuclear genes (Brown et al. 1979). As such, animal mtDNA has found widespread use as the molecular tool of choice for tracing matrilineal lineages (Cann et al. 1987; Hall and Muralidharan 1989; Sajantilla et al. 1996) and investigating shallow phylogenetic relationships, including intraspecific population structure, divergence time estimates, and the topology of relatedness between recently diverged species (Avise 1994, 2000). In contrast to metazoan mtDNA, plant mtDNA and chloroplast genomes generally represent the other end of the spectrum with extremely low rates of sequence evolution relative to the nuclear genome (Wolfe et al. 1987; Palmer and Herbon 1988; Gaut 1998; Cho et al. 2004); hence their use as molecular markers for dating deeper evolutionary events at higher taxonomic levels such as the origin of angiosperms (Bowe et al. 2000) and the monocot–dicot divergence (Chaw et al. 2004). Inheritance of mtDNA has long been assumed to be strictly uniparental through the maternal germline. This paradigm has been countered by accumulating evidence that biparental inheritance of mtDNA does occur via paternal leakage in both animal and plant taxa, albeit at a low frequency (Neale et al. 1989; Kondo et al. 1990; Gyllensten et al. 1991;

Kvist et al. 2003; Ballard and Whitlock 2004; Barr et al. 2005; McCauley et al. 2005; White et al. 2008). Unlike nuclear genomes, animal mitochondria are inherited in multiple copies (as a "population" of mitochondria, and with potentially more than one complete mitochondrial genome per mitochondrion) in the oocyte (Ballard and Whitlock 2004; Aanen et al. 2014). The population dynamics of mitochondria can be described as repeated population expansion-bottleneck cycles experienced every cell cycle of the host organism, whether mitotic (replicative segregation) or meiotic (Wallace and Chalkia 2013). Due to the high mitochondrial mutation rates in metazoans, cells may contain a number of genetically different mitochondria, usually referred to as a heteroplasmic population of mitochondria (Avise et al. 1987; Rand 2001; Ballard and Whitlock 2004). Heteroplasmy is generally assumed to be the result of spontaneous mutation within the maternal line (Avise et al. 1987). Furthermore, unlike the nuclear genome, animal mitochondria are thought to undergo limited or no recombination which manifests as high or complete linkage between mitochondrial loci (Moritz et al. 1987; but see Eyre-Walker et al. 1999).

Given the focal role of mutation in the origin of genetic variance and the rate of adaptation and divergence in evolution, several studies have investigated the rate and spectrum of spontaneous mitochondrial mutations. Furthermore, it had been posited that the conjunction of a high mutation rate with minimal recombination would predispose animal mitochondria to mutational meltdown due to the accumulation of a deleterious load of mutations via the process referred to as Muller's Ratchet (Lynch 1996; Lynch et al. 2006) although subsequent studies suggest that minimal levels of recombination could effectively counter mutation accumulation (Rand 2008; Neiman and Taylor 2009). However, recent studies have found no evidence for a disproportionately greater accumulation of deleterious mutations in the mtDNA relative to the nucleus (Cooper et al. 2015; Havird and Sloan 2016). Mutation rates can be inferred indirectly from analysis of sequence divergence between species. However, this approach suffers from the drawback of underestimation of the mutation rate, given that only mutation with minor or no fitness effects (nearly neutral or neutral mutations) tend to become fixed. A superior approach to studying the occurrence of mutations is the use of time- and labor-intensive mutation accumulation (MA) lines. MA experiments have been conducted in a handful of model organisms with life-history features and short generation times that allow for manipulation under stringent laboratory conditions (Halligan and Keightley 2009). In a typical MA experiment, multiple replicate lines descended from a single selfing founder or a pair of outcrossing founders is maintained independently for several hundred generations under extreme bottlenecking conditions to simultaneously attenuate the strength of natural selection and enable the fixation of originating mutations via random genetic drift. MA lines therefore facilitate the study of mutations before their purging from the genome via selection.

The *C. elegans* mitochondrial genome is typical of animal mitochondrial genomes with respect to size (13,794 bp) and gene content (total 36 genes comprising 12 protein-coding, 2 rRNA and 22 tRNA genes) (Okimoto et al. 1992; Lemire 2005). The mtDNA genomes of *C. elegans* MA lines were the first to be investigated for their rate and spectrum of spontaneous mutations (Denver et al. 2000). The authors employed Sanger sequencing that targeted ~76% of the mtDNA genome, but precluded a robust identification of heteroplasmic variants at low frequencies. A similar approach was used to identify mtDNA variants in the congener species *C. briggsae* (Howe et al. 2010), and the microcrustacean *Daphnia pulex* (Xu et al. 2012). The advent of massively parallel sequencing technologies have facilitated direct sequencing of MA lines to generate direct genome-wide estimates of the rate and spectrum of spontaneous mitochondrial mutations in an additional five unicellular/multicellular eukaryote species, namely, *Drosophila melanogaster* (Haag-Liautard et al. 2008), *Saccharomyces cerevisiae* (Lynch et al. 2008), the nematode *Pristionchus pacificus* (Molnar et al. 2011), and the protists *Paramecium tetraurelia* (Sung et al. 2012) and *Dictyostelium discoideum* (Saxer et al. 2012). The overall rate of mtDNA spontaneous mutation per site per generation gleaned from spontaneous MA experiments appears fairly consistent across diverse unicellular and multicellular eukaryotic species, ranging approximately 2-fold from 0.76 to 1.6×10^{-7}. In contrast, the reported mutational spectrum of base substitutions in the mtDNA genome is markedly divergent across different taxa (Montooth and Rand 2008). Metazoan mitochondrial genomes are significantly A + T-biased with respect to base composition. A strongly biased mutation pressure towards A/T changes has been reported for *D. melanogaster* (Haag-Liautard et al. 2008), and the two nematode species, *C. briggsae* (Howe et al. 2010) and *P. pacificus* (Molnar et al. 2011). However, a G/C mutation bias has been reported for two species, namely *C. elegans* (Denver et al. 2000) and *S. cerevisiae* (Lynch et al. 2008).

The loss or fixation of mutations and their consequences for population fitness depend upon the selection coefficients associated with individual mutations and the effective population size, N_e. Preceding spontaneous MA studies in this classical field of enquiry have maintained the focal organism at a constant minimal population size to generate species-specific mutation parameters such as the spontaneous mutation rate and the average fitness effect of mutations (reviewed in Halligan and Keightley 2009). We created parallel experimental evolution lines of *C. elegans* at varying population sizes representing the first long-term spontaneous MA experiment of its kind for any species by consecutively bottlenecking 35 lines at population sizes of $N = 1$, 10, and 100 individuals for 409 generations (Katju et al. 2015). Using massively parallel Illumina paired-ends sequencing technology in combination with our long-term MA experiment, we provide the first comprehensive analysis of the genome-wide distribution of spontaneous mitochondrial mutations under differing intensity of selection. The varying population size treatment is a significant advance in the experimental design of MA studies and offers a powerful framework to investigate how spontaneous mutational input in conjunction with varying strengths of natural selections shapes genetic variation and levels of mitochondrial heteroplasmy in a model

eukaryote. Moreover, we compare the spectrum of spontaneous mtDNA mutations in our experimental lines with those observed in the genomes of 38 *C. elegans* natural isolates. It had been previously concluded that the *C. elegans* mitochondrial genome has a strong A/T → G/C mutation bias, and therefore that the extant G + C content of the genome was due to natural selection and not mutation pressure (Denver et al. 2000). However, we find evidence for a strong G/C → A/T mutation bias which is concordant with results in other metazoans as well as *C. briggsae* (Howe et al. 2010). In addition, we find evidence for pervasive heteroplasmy in the *C. elegans* mitochondrial genome which also conflicts with previous conclusions in *C. elegans* (Denver et al. 2000) but is consistent with subsequent results in *C. briggsae* (Howe et al. 2010). We provide the first maximum-likelihood estimate of the mitochondrial effective populations size, $N_{e[mtDNA]}$, for *C. elegans*, which exceeds similar estimates for two other multicellular eukaryotes (Haag-Liautard et al. 2008; Xu et al. 2012). Finally, we demonstrate for the first time, in a stringent experimental setting, that the accumulated frequency of nonsynonymous and frameshift mutations in the mitochondrial genome declines with increasing population size. Although this relationship is predicted from population-genetic theory, and supported by observations in natural populations, this is the first time that this relationship has been demonstrated with spontaneous mutations in MA experiments.

Results

Experimental Overview

We sequenced 86 near-complete mitochondrial genomes of *C. elegans* lines and their *N2* ancestor comprising a long-term MA experiment spanning 409 MA generations with differing population sizes of $N = 1$, 10, 100 (fig. 1). For the 20 MA lines (1A–1T) maintained at population size $N = 1$ and the ancestral pre-MA *N2* control, the mtDNA genome of a population of worms derived from one hermaphrodite per line was sequenced. For the 20 $N = 1$ lines, a total of 268,256 bp were surveyed which represented an average 97.2% (13,413 bp) of the 13,794-bp *C. elegans* mitochondrial genome. The remaining 381 bp encompass the D-loop, which contains the origin of replication and is characterized by A + T-rich, low-complexity sequence tracts. Because lines comprising larger population sizes are expected to harbor greater standing genetic variation, the genomes of four and five individuals were sequenced per $N = 10$ (10 lines; 10A–10J) and $N = 100$ (five lines; 100A–100E) line, respectively. This sequencing design yielded 40 and 25 mtDNA genomes for the $N = 10$ and $N = 100$ MA lines, respectively. The average read depth of the mitochondrial genomes was 388×, 187×, and 197× per genome within the population size treatments of $N = 1$, 10, and 100, respectively (supplementary table S1, Supplementary Material online). A total of 46 variants were detected among all sequenced MA lines; 24 variants in 20 $N = 1$ MA lines, 13 variants in 10 $N = 10$ MA lines, and nine variants in five $N = 100$ MA lines. With respect to the class of mutations, we detected: (i) 11 base substitutions, (ii) 32 small insertions or deletions (1–3 bp) in homopolymeric runs, and (iii) three large deletions ranging in size from 107 to 1,185 bp (tables 1–3). Because differing intensities of selection *vs.* drift were hypothesized for the three different population sizes, we analyzed the mutation rates and spectrum separately for each population size treatment.

Spontaneous Mitochondrial Mutation Rates and Spectrum in *C. elegans* from $N = 1$ Lines

Each line in our $N = 1$ population size treatment was propagated by a single random hermaphrodite, resulting in an effective population size of $N_e = 1$. Under this bottlenecking regimen, natural selection can only purge the most lethal mutations (causing sterility and/or mortality) whereas the effects of genetic drift are maximized. Genetic drift renders many strongly deleterious mutations effectively neutral, enabling their persistence in the genome. Sequence analysis of the 20 $N = 1$ genomes maintained for an average ~363 MA generations detected 24 mutations (table 1), yielding an overall spontaneous mitochondrial mutation rate, μ_{total}, of 1.05×10^{-7} per site per generation (95% CI: 0.52 to 1.80×10^{-7}).

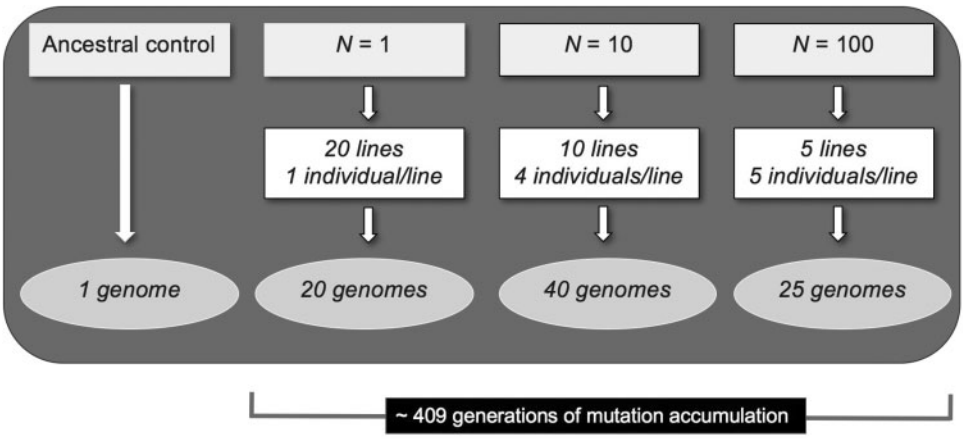

FIG. 1. Whole-genome sequencing to yield 86 *C. elegans* genomes, including that of the ancestral, pre-mutation accumulation *N2* control and 35 MA lines following ~409 MA generations. Multiple individuals were sequenced for MA lines maintained at larger population sizes ($N = 10$ and 100).

Table 1. List and Details of the 24 mtDNA Variants Detected in 20 *C. elegans* MA Lines Maintained at Population Size $N = 1$ for an Average 363 MA Generations.

MA Line	Generation	Position	Mutation	Effect	Genomic Region	Frequency	Context
1O	397	712	C→A	Ser→Stop	ND4L	0.49	TCTGTTATTTCAAGAATCCTG
1P	393	732	G→T	Val→Phe	ND4L	0.14	GGGTATGGTAGTTATAGTAGG
1G	346	1679	G→T	tRNA	tRNA-Asn	0.13	ATTGTTGACTGTTAATCAAAA
1T	305	2040	$(T)_7$→$(T)_8$	Frameshift	ND1	0.19	CGTTACCATA[T]$_7$GATTTTATTA
1P	393	3234	T→A	-	Noncoding	0.04	TAATGAGTAATAAAAAAAAAA
1M	395	3235	$(A)_{10}$→$(A)_9$	-	Noncoding	0.03	TAATGAGTAAT[A]$_{10}$GATGTTAACT
1D	326	4475	C→T	tRNA	tRNA-Phe	0.88	TAAAATATGGCCCTGAAGAGG
1G	346	4809–5307	499-bp del[a]	Frameshift	ctb-1	0.96	TATTTTTTAT[AAG...TGA]TATTTTTAT
1I	368	6699	$(A)_7$→$(A)_6$	Frameshift	ND4	0.18	GTTATTAGAG[A]$_7$TAATAATTTA
1H	281	8654	$(A)_7$→$(A)_8$	Frameshift	COI	0.03	ATTTAACAGG[A]$_7$GAAGTTTTTG
1C	392	9437–9543	107-bp del[a]	tRNA	tRNA-Cys, tRNA-Met, tRNA-Asp	0.76	CTTAGTATAA[TTT...GTT]TTAGTATAAA
1P	393	10100	C→T	Thr→Ile	COII	0.67	CCTTGTGATACTAACATTCGT
1C	392	11722	$(T)_8$→$(T)_9$	Frameshift	ND5	0.40	ATTGGATTTG[T]$_8$ATAGGTGGAA
1K	349	11722	$(T)_8$→$(T)_9$	Frameshift	ND5	0.42	ATTGGATTTG[T]$_8$ATAGGTGGAA
1L	342	11722	$(T)_8$→$(T)_9$	Frameshift	ND5	0.58	ATTGGATTTG[T]$_8$ATAGGTGGAA
1L	342	11722	$(T)_8$→$(T)_{10}$	Frameshift	ND5	0.03	ATTGGATTTG[T]$_8$ATAGGTGGAA
1R	386	11722	$(T)_8$→$(T)_9$	Frameshift	ND5	0.78	ATTGGATTTG[T]$_8$ATAGGTGGAA
1S	291	11722	$(T)_8$→$(T)_9$	Frameshift	ND5	0.06	ATTGGATTTG[T]$_8$ATAGGTGGAA
1G	346	11778	$(T)_8$→$(T)_9$	Frameshift	ND5	0.03	ATTAGGAATC[T]$_8$AGAATGAGAT
1G	346	11778	$(T)_8$→$(T)_{10}$	Frameshift	ND5	0.70	ATTAGGAATC[T]$_8$AGAATGAGAT
1H	281	11778	$(T)_8$→$(T)_7$	Frameshift	ND5	0.82	ATTAGGAATC[T]$_8$AGAATGAGAT
1G	346	12304	C→T	Thr→Ile	ND5	0.94	CATAGTAGAACTTTAGTTACA
1D	326	12580	G→A	Gly→Asp	ND5	0.88	ATACAAGTAGGTTACATTATT
1I	368	12965	C→T	Synonymous	ND5	0.03	TTCTTAACATCCCAAGACTTT

NOTE.—The mutations are displayed as changes on the major strand.
[a]del = deletion.

Table 2. List and Details of the 13 mtDNA Variants Detected in 10 *C. elegans* MA Lines Maintained at Population Size $N = 10$ for an Average 408.6 MA Generations.

MA Line	Generation	Position	Mutation	Effect	Genomic Region	Frequency	Context
[1]10B	409	190	$(T)_7$→$(T)_6$	Frameshift	ND6	0.33	TAAAAAGAAG[T]$_7$CTTATCTTTT
[2]10C	408	2078	$(T)_8$→$(T)_9$	Frameshift	ND1	0.02	TTCAGTTTTA[T]$_8$ATGTTTAATT
[3]10A	409	3235	$(A)_{10}$→$(A)_9$	–	Noncoding	0.04	TAATGAGTAAT[A]$_{10}$GATGTTAACT
[4]10I	409	3235	$(A)_{10}$→$(A)_{13}$	–	Noncoding	0.10	TAATGAGTAAT[A]$_{10}$GATGTTAACT
[5]10H	408	6699	$(A)_7$→$(A)_6$	Frameshift	ND4	0.11	GTTATTAGAG[A]$_7$TAATAATTTA
[6]10F	409	6753	$(T)_8$→$(T)_9$	Frameshift	ND4	0.44	ATTTGTATTA[T]$_8$ATTCCTAGTA
[7]10C	408	7920–9104	1185 bp del[a]	In-frame deletion	COI	0.82	TATCGGAACT[CTT...GTA]TTATTTTATT
[8]10E	409	9723	T→C	Synonymous	COII	1.00	ATTGATTTCATAGGTTTAATT
[9]10B	409	11704	$(T)_7$→$(T)_8$	Frameshift	ND5	0.47	AATATTTCTA[T]$_7$GATTGGATTT
[10]10B	409	11722	$(T)_8$→$(T)_9$	Frameshift	ND5	0.45	ATTGGATTTG[T]$_8$ATAGGTGGAA
[11]10D	408	11722	$(T)_8$→$(T)_9$	Frameshift	ND5	0.04	ATTGGATTTG[T]$_8$ATAGGTGGAA
[12]10D	408	11778	$(T)_8$→$(T)_7$	Frameshift	ND5	0.02	ATTAGGAATC[T]$_8$AGAATGAGAT
[13]10F	409	13255	$(T)_8$→$(T)_7$	Frameshift	ND5	0.02	TTAATTTTTA[T]$_8$ATAATTTGTT

NOTE.—The mutations are displayed as changes on the major strand. The reported frequency of each mutation is an average of the frequencies detected in the four individuals sequenced from each of the $N = 10$ lines.
[a]del = deletion.
Individual frequencies of the mutation in each of the four individuals sequenced per line: [1]0.2, 0.31, 0.36, 0.46; [2]0.01, 0.01, 0.02, 0.03; [3]0.03, 0.03, 0.04, 0.05; [4]0.11, 0.11, 0.11, 0.05; [5]0, 0.06, 0.087, 0.29; [6]0.34, 0.38, 0.43, 0.61; [7]0.77, 0.85, 0.79, 0.81; [8]0.99, 0.99, 1.0, 1.0; [9]0.38, 0.47, 0.49, 0.53; [10]0.38, 0.4, 0.5, 0.52; [11]0, 0.02, 0.03, 0.11; [12]0, 0.01, 0.01, 0.04; [13]0.01, 0.01, 0.01, 0.03.

The $N = 1$ MA lines contained nine base substitutions in frequencies ranging from 3% to near fixation (table 1), yielding a spontaneous rate of base substitution, μ_{bs}, of 4.32×10^{-8} per nucleotide site per generation (95% CI: 0.87 to 9.44×10^{-8}). 67% (6/9) of these base substitutions occurred in three of the 12 mitochondrial protein-coding genes (ND5, three substitutions; ND4L, 2 substitutions; COII, one substitution), and five of these were nonsynonymous substitutions. Twenty-two percent (2/9) of the base substitutions occurred in tRNA genes (tRNA-Asn and tRNA-Phe) and a single base substitution occurred in the noncoding region between ATP6 and tRNA-Lys. Eight of the nine base substitutions were from G/C→A/T and only one substitution was in the converse direction, from A/T→G/C. The observed spectrum of base

Table 3. List and Details of the Nine mtDNA Variants Detected in Five *C. elegans* MA Lines Maintained at Population Size $N = 100$ for an Average 409 MA Generations.

MA Line	Generation	Position	Mutation	Effect	Genomic Region	Frequency	Context
[1]100B	409	2078	$(T)_8 \rightarrow (T)_9$	Frameshift	ND1	0.01	TTCAGTTTTA[T]$_8$ATGTTTAATT
[2]100B	409	3235	$(A)_{10} \rightarrow (A)_9$	–	Noncoding	0.02	TAATGAGTAAT[A]$_{10}$GATGTTAACT
[3]100C	409	3235	$(A)_{10} \rightarrow (A)_8$	–	Noncoding	0.80	TAATGAGTAAT[A]$_{10}$GATGTTAACT
[4]100D	409	3235	$(A)_{10} \rightarrow (A)_9$	–	Noncoding	0.04	TAATGAGTAAT[A]$_{10}$GATGTTAACT
[5]100A	409	6753	$(T)_8 \rightarrow (T)_9$	Frameshift	ND4	0.20	ATTTGTATTA[T]$_8$ATTCCTAGTA
[6]100C	409	11722	$(T)_8 \rightarrow (T)_9$	Frameshift	ND5	0.10	ATTGGATTTG[T]$_8$ATAGGTGGAA
[7]100C	409	11778	$(T)_8 \rightarrow (T)_9$	Frameshift	ND5	0.04	ATTAGGAATC[T]$_8$AGAATGAGAT
[8]100D	409	11778	$(T)_8 \rightarrow (T)_7$	Frameshift	ND5	0.01	ATTAGGAATC[T]$_8$AGAATGAGAT
[9]100E	409	12147	G→A	Gly→Ser	ND5	0.59	GGTTTTTAGAGGTTATTATTT

NOTE.—The mutations are displayed as changes on the major strand. The reported frequency of each mutation is an average of the frequencies detected in the five individuals sequenced from each of the $N = 100$ lines. Individual frequencies of the mutation in each of the five individuals sequenced per line:
[1]0, 0, 0, 0, 0.05; [2]0.05, 0, 0.01, 0.01, 0.05; [3]0.84, 0.77, 0.79, 0.78, 0.81; [4]0.06, 0.04, 0.06, 0.03, 0.02; [5]0, 0.01, 0.02, 0.45, 0.52; [6]0, 0.03, 0.06, 0.19, 0.21; [7]0, 0, 0, 0.01, 0.21; [8]0, 0, 0.01, 0.01, 0.03; [9]0, 0, 0.98, 0.99, 1.0.

substitutions in our study is significantly different from a previous study of the mutation rate in the mitochondria of *C. elegans* MA lines (fig. 2) which reported four G/C→A/T and 10 A/T→G/C mutations (Denver et al. 2000) (two-tailed Fisher's Exact Test, $P = 0.0094$). However, we do note that our patterns of observed mtDNA base substitutions are almost identical to preceding studies of *D. melanogaster* (Haag-Liautard et al. 2008) (two-tailed Fisher's Exact Test, $P = 1$) (fig. 2) and *C. briggsae* (Howe et al. 2010) (two-tailed Fisher's Exact Test, $P = 1$) MA lines.

The total A + T content of the *C. elegans* mitochondrial genome is 75.5%, and the A + T bias is more pronounced in the third codon positions (86%) and intergenic regions (93%) (Okimoto et al. 1992), which are less likely to be under selection than the first two codon positions and thus, more likely to reflect the cumulative effects of prevailing mutational biases in the genome. The equilibrium A + T content is expected to equal $u/(u + v)$ where u is the mutation rate from a G/C to an A/T pair and v is the mutation rate from an A/T to a G/C pair (Sueoka 1962). If we assume that the 86% A + T represents the equilibrium A + T content in the absence of selection, the mutation rate from G/C→A/T should be ~six-fold higher than the mutation rate from A/T→G/C. Given an approximately 75% A + T composition of the *C. elegans* mtDNA genome and assuming a 6-fold higher rate of G/C→A/T mutations relative to A/T→G/C ones, we would only expect twice as many G/C→A/T than A/T→G/C mutations. Our observed 8:1 ratio of G/C→A/T substitutions is not significantly different from this expectation (two-tailed Fisher's Exact Test, $P = 0.58$).

We further investigated heterogeneity in the mutation rate of nucleotides by examining the frequency of transitions versus transversions. Of the nine base substitutions observed in the $N = 1$ lines, five and four were transitions and transversions, respectively. If there are equal substitution rates between nucleotides, the ratio of transitions:transversions (Ts:Tv henceforth) is expected to be 1:2. While we lack sufficient power to test for statistical deviations of our observed Ts:Tv ratio from this null model, we observe a two-fold excess in the occurrence of transitions. However, our transition bias is not as strong as the 25:3 or 13:3 Ts:Tv bias observed in *D. melanogaster* (Haag-Liautard et al. 2008) and other *C. elegans* MA lines (Denver et al. 2000), respectively.

Insertion/deletion (indel) events represented a larger fraction of mutational events observed in the 20 $N = 1$ lines (15/24; 62.5%), comprising 10 insertions and five deletions (table 1) and yielding a spontaneous indel rate of 6.14×10^{-8} per site per generation (95% CI: $2.56–11.1 \times 10^{-8}$). Two of these are large deletions of 107 and 499 bp, respectively (table 1). Line 1C had a 107 bp deletion (frequency 0.76 in the population) that spanned part of the tRNA-Cys, the entire tRNA-Met and the 5′ end of the tRNA-Asp genes. The 499-bp deletion in the *ctb-1* gene in MA line 1G is likely deleterious in fitness. This mtDNA variant is near fixation in the population (frequency 0.96) and it shortens a 370 aa encoded protein by 166 aa and simultaneously results in a frameshift mutation in the undeleted downstream amino acids. The relative mean productivity for line 1G following 346 MA generations was 0.70 (ancestral control was assigned a value of 1) (Katju et al. 2015). The relative contributions of the nuclear versus mitochondrial genetic load to this fitness decline cannot be disentangled from this data alone. Likewise, line 1C displayed a decline in relative mean productivity (0.76) at 392 MA generations (Katju et al. 2015). Excluding these two large deletions, the remaining 13 of these 15 mutations are small indels (1–3 bp) in homopolymeric runs of A or T nucleotides. Ten of the 13 small indels are insertions with a mean size of 1.2 bp. The remaining three of the 13 small indels are 1-bp deletions.

Distribution of Spontaneous mtDNA Mutations at Larger Population Sizes ($N = 10$ and 100)

Eight of the $N = 10$ lines (total 10 lines, 10A–10J) harbored 13 mitochondrial mutations (table 2). Two lines, 10G and 10J did not accumulate any observable mitochondrial mutations through the course of the 409 generation experiment. Because mutations in the $N = 10$ or 100 lines were always found in multiple individuals sequenced from within the same line, we considered this to be a confirmation of the mutations. Therefore, if at least one individual in a particular $N = 10$ or 100 line possessed a heteroplasmic mutation that reached a 3% threshold in frequency but the same variant was detected in <3% frequency in other individuals of the same

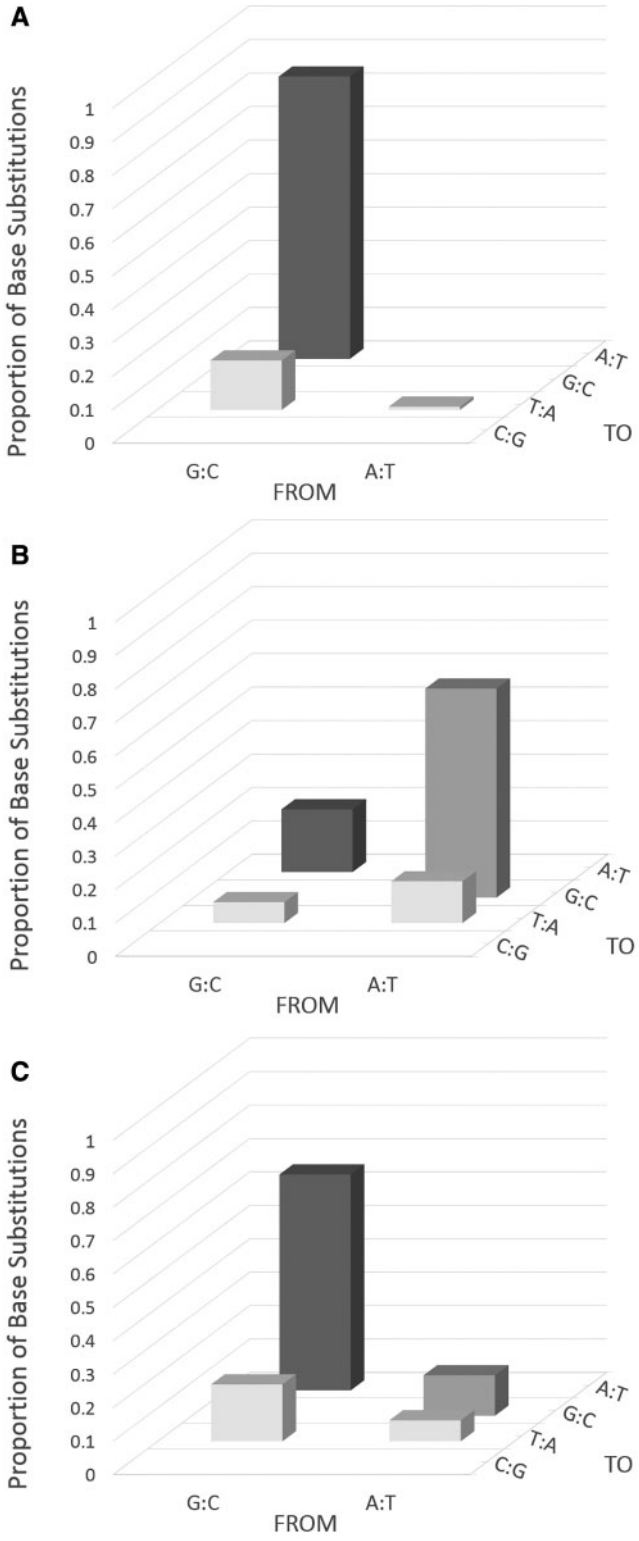

FIG. 2. Mutation spectrum of base substitutions in the mitochondrial genome. Transitions are represented in dark and medium gray, while the four possible transversions are in light gray and black. The proportion of base substitutions are normalized by the frequencies of each individual variant. (*A*) The bias in spontaneous base substitutions detected in our study of 20 *C. elegans* $N = 1$ MA lines is contrasted with the spectrum of mtDNA base substitutions observed in preceding studies of (*B*) *C. elegans* MA lines (Denver et al. 2000) and (*C*) *C. briggsae* MA lines (Howe et al. 2010).

line, we calculated and reported the average frequency of the variants across all individuals sequenced from within the same MA line. Hence, some of the variant frequency values are <3% in tables 2 and 3. The majority of the observed mutations (11 of 13) in the $N = 10$ lines were single nucleotide indels in homopolymeric runs with frequencies ranging from 0.02 to 0.47. The two remaining mutations in the $N = 10$ lines were a single base substitution and one large deletion. Line 10E contained one synonymous base substitution (CAT→CAC) in the COII gene which had reached fixation. Line 10C had a large 1,185 bp in-frame deletion within the COI gene that had reached a frequency of 0.82 in the population. The observed mitochondrial mutation rate for the $N = 10$ populations was 6.95×10^{-8}/site/generation (table 4).

The five lines within $N = 100$ population size treatment (100A–100E) harbored nine mitochondrial mutations (table 3). As was observed in the $N = 10$ lines, the vast majority of the mutations (eight of nine) in the $N = 100$ are single nucleotide indels in homopolymeric runs, ranging in frequencies from 0.01 to 0.80. Only one base substitution was observed within the $N = 100$ experimental treatment, namely a nonsynonymous change (GGA→AGA) in the ND5 gene with a population frequency of 0.59. The observed mitochondrial mutation rate for the $N = 100$ populations was 6.47×10^{-8}/site/generation (table 4).

Differences in Frequency of mtDNA Mutations with Varying Intensities of Natural Selection

Once mutations have arisen, how do they fare within lines across different population sizes? For each sequenced individual, we calculated the sum of the frequency of variants exceeding 3% divided by the number of nucleotides sequenced and the number of generations. Individuals lacking variants that reached the ≥3% criteria for detection were not included. For the larger population sizes, we calculated the average sum of the frequency of variants per line using only those individuals with a variant frequency of ≥ 3% as was done in the case of the $N = 1$ lines. This is similar to the calculation of mutation rate per line except that only individuals with mutations were included. The average frequency of variants per MA line was compared across the three population size treatments ($N = 1$, 10 and 100 with sample sizes $n = 13, 8, 5$ MA lines, respectively) (fig. 3). We did not observe a statistically significant difference in the average frequency of variants per MA line between the three population size treatments (parametric, ANOVA, F-ratio = 1.65, $df = 2$, $P = 0.215$; nonparametric, Kruskal–Wallis, $df = 2$, $P = 0.294$).

Most detrimental mitochondrial mutations need to reach a critical threshold in frequency of affected molecules before they have a noticeable effect on the phenotype (Stewart and Chinnery 2015). Can selection suppress the frequencies of nonsynonymous/indel mutations below the hypothesized critical threshold within individuals in the larger-sized populations ($N = 10$, 100)? We compared the frequencies of mutations that are likely to be subject to purifying selection (nonsynonymous mutations in coding regions and indels within genic regions including RNA genes) across the three

Table 4. Estimates of per Site per Generation Mutation Rates and the Mitochondrial Effective Population Size ($N_{e[mtDNA]}$) per Worm Generation for the Mitochondrial Genome.

Mutation Class	Population Size	Number of Mutations	$\mu \times 10^{-8}$ [95% CI]	$N_{e[mtDNA]}$ (ML) [95% CI]
All events	$N = 1$	24	10.46 [5.2, 18]	62 [33, 117]
	$N = 10$	10	6.95	
	$N = 100$	6	6.47	
Base Substitutions	$N = 1$	9	4.32 [0.87, 9.44]	60 [22, 146]
	$N = 10$	1	1.83	
	$N = 100$	1	2.16	
Indels	$N = 1$	15	6.14 [2.56, 11.1]	
	$N = 10$	9	5.12	
	$N = 100$	8	4.31	

NOTE.—Mutation rate estimates for $N = 1$ lines represent the spontaneous rate of origin of mtDNA mutations with minimal influence of selection. Rate estimates for $N = 10$ and $N = 100$ represent substitution patterns with increasing intensity of natural selection. Mutations with frequencies below 3% are not included in the rate estimates.

population sizes within our MA experiment. The average frequency of nonsynonymous and indel mutations showed a significant negative correlation with population size ($r = -0.51$, $P = 0.019$) (fig. 4A). To test if there is an effect of differences in read depth per genome, we subsampled the $N = 1$ reads to create a new data set with read depths similar to the average read depth of $N = 10$ and 100 individuals. The analysis described above was repeated with the reduced $N = 1$ datasets to test if differences in read depth contributed to the results. Despite the loss of some heteroplasmic variants due to the change in average read depth, the random assignment of reads into the new dataset also led to increased frequencies of some other heteroplasmic variants. There was no qualitative change in the correlation using the reduced $N = 1$ dataset and the association between the accumulation of nonsynonymous and indel mutations with population size remained significant ($r = -0.60$, $P = 0.010$). In contrast, the frequency of synonymous and intergenic variants per individual was not correlated with population size ($r = 0.1$, $P = 0.77$) (fig. 4B). This suggests that selection is indeed impeding the accumulation of nonsynonymous mutations at larger population sizes ($N = 10$ and 100).

Observed mtDNA Polymorphisms in Natural Isolates of *C. elegans*

The variation in mitochondrial genomes was also assessed in 38 natural isolates (supplementary table S2, Supplementary Material online) whose genomes had been previously sequenced (Thompson et al. 2013). A total of 426 mtDNA variants were identified in these 38 natural isolates of which 389 (91.3%) and 37 (8.7%) were polymorphic (>80% within individuals) and heteroplasmic (frequency ≤ 80%), respectively (supplementary table S3, Supplementary Material online). Most variants detected in the natural isolates are either very high- or very low-frequency variants (fig. 5), with proportionally fewer variants at intermediate frequencies. This pattern contrasts with the variants identified in the MA lines.

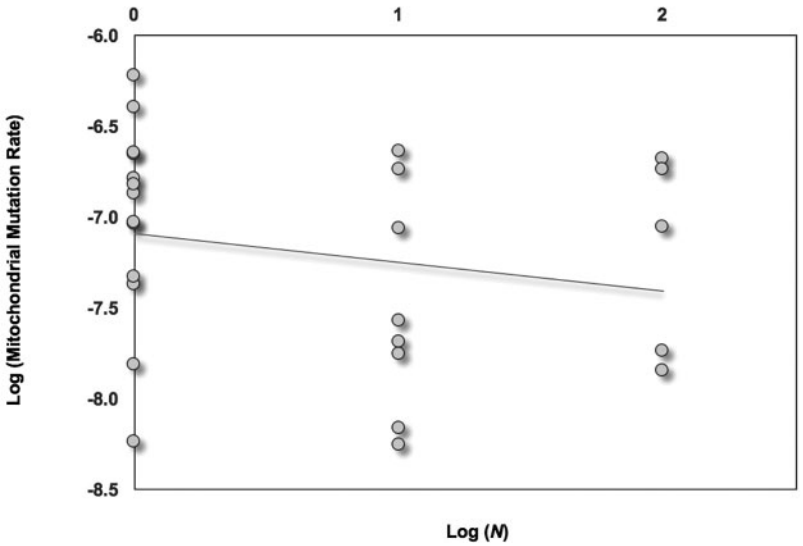

FIG. 3. Distribution of observed overall mitochondrial mutation rates (all events including base substitutions and indels) for individual MA lines within three population size treatments ($N = 1$, 10, and 100 hermaphrodites). Observed rates were calculated per site per generation. Individual MA lines with zero mutations were excluded from the analysis resulting in sample sizes of 13 $N = 1$ lines, eight $N = 10$ lines and five $N = 100$ lines. Values are presented in log scale.

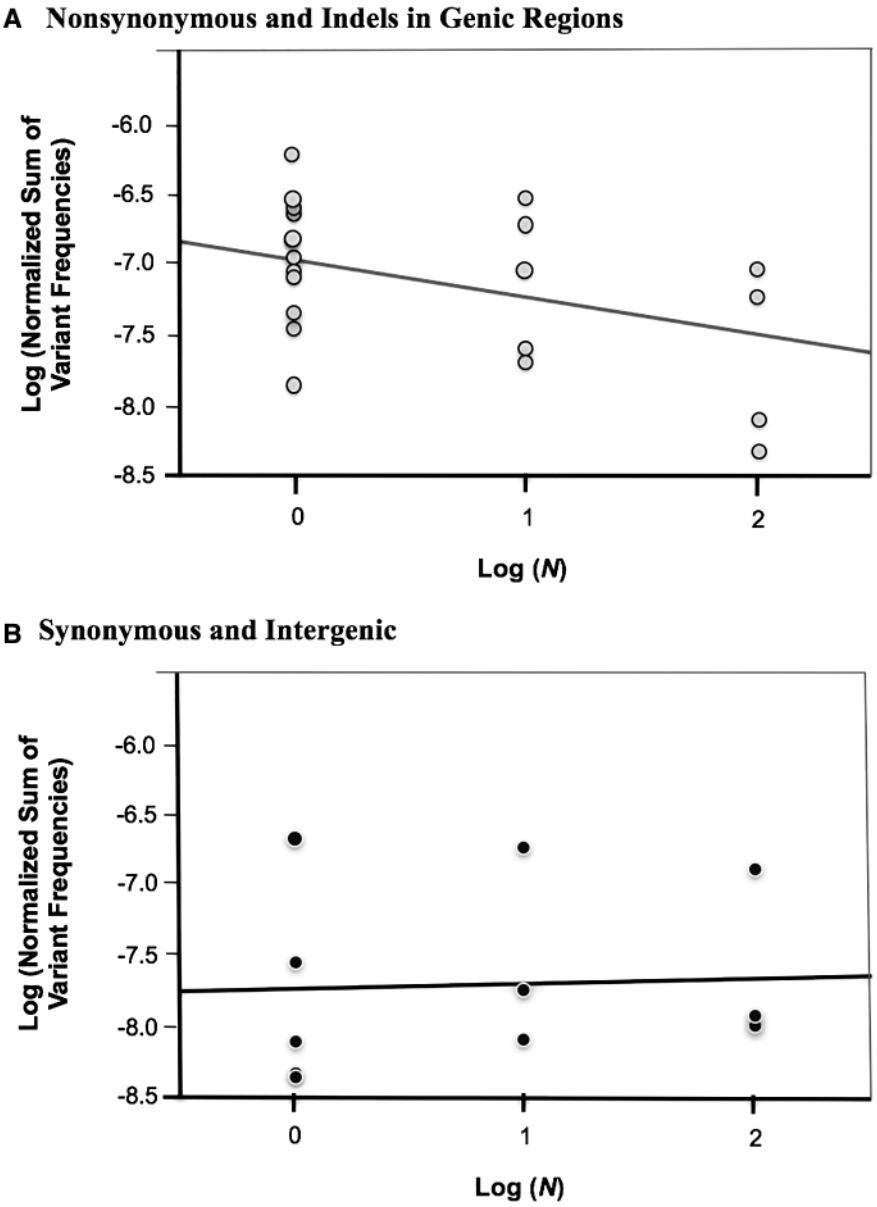

FIG. 4. The relationship between the normalized sum of the frequency of variants within individual MA lines and the population size. Values are presented in log scale. MA lines with zero mutations were excluded from this analysis. The sum of the frequency of variants was normalized by the number of base pairs analyzed for each line, and the number of generations. (*A*) The normalized sum of nonsynonymous, large deletion, and frameshift mutations in genic regions that may be under selection. (*B*) The normalized sum of putative neutral mutations comprising synonymous substitutions and intergenic mutations.

Of the 408 base substitutions identified, 387 were polymorphic (~95%) and 21 were heteroplasmic (~5%). Nonsynonymous SNPs constituted ~17% of 357 polymorphic and 55% of 20 heteroplasmic coding-region SNPs, respectively. The low frequency of nonsynonymous SNPs in *C. elegans* natural isolates also contrasts with the relatively high ratio of nonsynonymous mutations in the MA lines (fig. 6, two-tailed Fisher's Exact Test, $P = 0.0003$), consistent with the expectation that nonsynonymous substitutions are under purifying selection.

The overall Ts/Tv ratio of polymorphic SNPs is 5.5. The Ts/Tv ratio is 3.3, 3.7, and 25.8 at nondegenerate, four-fold, and two-fold degenerate sites, respectively. The high Ts/Tv ratio at two-fold degenerate sites is expected, because all transversions at these sites are nonsynonymous and will likely be selected against, whereas all transitions are synonymous and more likely to be neutral. Because the consequences of transitions and transversions at either four-fold (all synonymous) or nondegenerate (all nonsynonymous) sites do not differ, the average Ts/Tv ratio of 3.5 at these sites (nondegenerate and four-fold degenerate) should reflect the Ts/Tv ratio of spontaneous mutations.

The number of mutations from G/C→A/T and A/T→G/C polymorphic (high-frequency) sites in protein-coding genes are approximately equal. However, G/C→A/T mutations are roughly 7.3× more frequent than A/T→G/C mutations at synonymous sites after adjusting for base composition bias. Using $u/(u+v)$ to calculate the expected equilibrium A/T

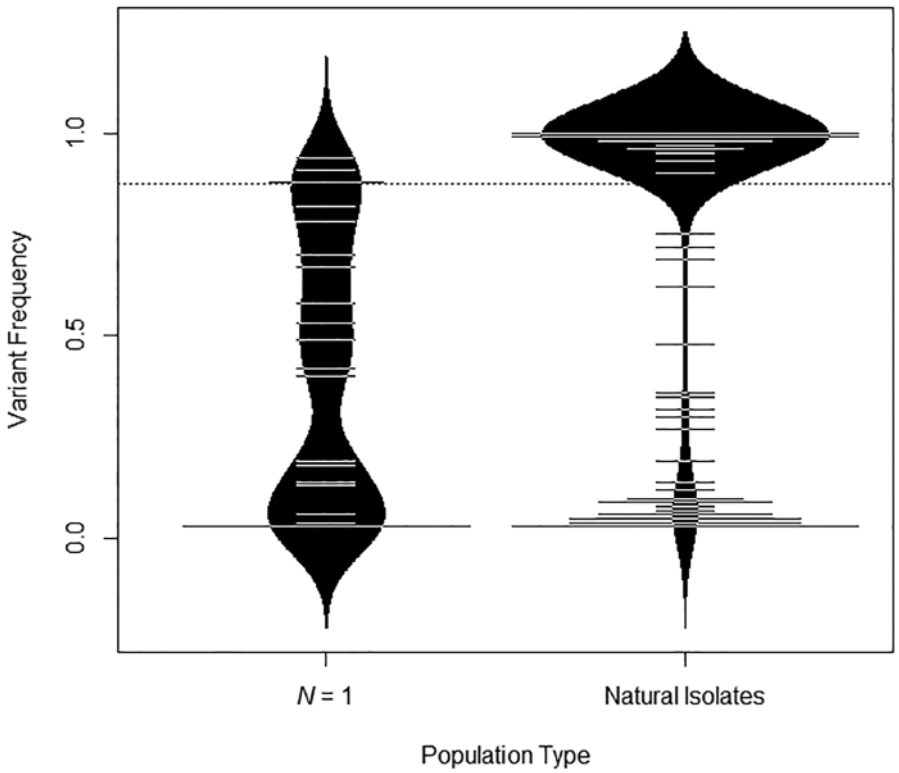

FIG. 5. The distribution of mtDNA variant frequencies within the $N = 1$ MA lines ($n = 20$) lines and 38 *C. elegans* natural isolates. Variant frequencies that fall below the dotted line (0.95) are heteroplasmic. The distribution of variants within the MA lines is approximately uniform whereas that of the natural isolates appears bimodal. The majority of variants detected within the natural isolates have reached fixation whereas the heteroplasmic variants tend to be in low frequency.

content at synonymous sites based on a 7.3× higher rate of G/C→A/T mutations than in the opposite direction yields 88% A/T, which is similar to the observed A/T content of 86% at third codon positions.

A total of 18 indels were identified in the 38 natural isolates. 16 indels were heteroplasmic (90%), all of which occurred at low frequencies ranging from 0.03 to 0.13. The two remaining indels were polymorphic, both of which were located in noncoding sequences. Thirteen indels (81%) caused frameshifts in protein-coding sequences and the majority of these (7/13) were in the protein-coding gene ND5. The high proportion of ND5 indels in natural isolates mirrors the results in the MA lines wherein 14 of 22 indels causing frameshifts were present in ND5. Whereas the ND5 indels in the MA lines were found in frequencies ranging from 3% to 94%, in natural isolates they exhibit a much narrower distribution of low frequencies (3–12%).

Mutational Hotspots

Certain sites in the *C. elegans* mitochondrial genome show parallel heteroplasmies in different lines. In the $N = 1$ lines, 12 mutations (50%) were detected in the ND5 gene. Nine of the 12 ND5 mutations are small indels occurring in two homopolymeric runs of eight thymines, separated by 48 bp. Six of these indels have reached reasonably high frequency (40-80%), whereas three remain in low frequency (3–6%). In addition to the indels, three substitutions were detected in ND5, two nonsynonymous mutations in high frequency

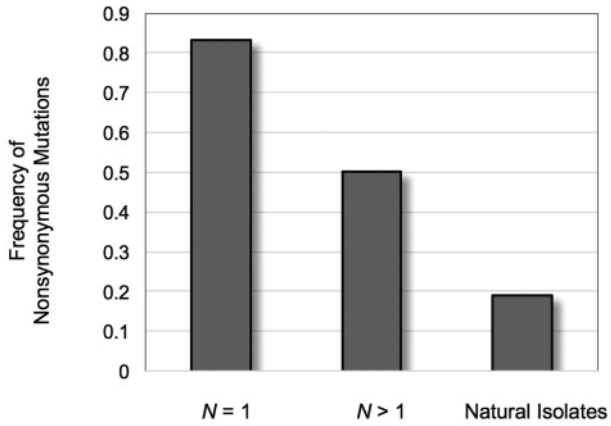

FIG. 6. Relative frequencies of nonsynonymous mtDNA mutations in *C. elegans* MA lines of varying population sizes and 38 natural isolates. The frequency of nonsynonymous changes is highest in the $N = 1$ MA lines ($n = 20$ lines). The $N > 1$ category ($n = 15$ lines) comprises ten $N = 10$ lines and five $N = 100$ lines. Lower relative frequencies of nonsynonymous mutations in the natural isolates is consistent with their deleterious consequences and eradication via natural selection.

(88 and 94%) and one synonymous mutation in low frequency (3%). The high proportion of ND5 mutations is also present in populations with larger bottlenecks, $N = 10$ and $N = 100$. In the $N = 10$ populations, five of the 12 mutations are in ND5, two in intermediate frequency and three in low frequency. All of these are small indels in three separate

homopolymeric runs. In $N = 100$, four of the nine mutations are indels in ND5. Short indels in homopolymeric runs were found in runs of seven or more nucleotides and the ND5 gene harbored 11 of the 33 homopolymeric runs within this range (fig. 7). The high proportion of homopolymeric runs in ND5 is a likely contributing factor to the high incidence of ND5 mutations in these experiments.

Effective Population Size of Mitochondria ($N_{e[mtDNA]}$) in *C. elegans*

A maximum likelihood (ML) procedure (Haag-Liautard et al. 2008) was employed to provide an estimate of the effective number of maternal mitochondrial genomes inherited by the progeny per generation, based on the distribution of frequencies of mutations within individuals. ML estimates of the mitochondrial effective population size ($N_{e[mtDNA]}$) using the observed frequencies of base substitutions, and the overall set of mutations (base substitutions and indels combined) suggested that ~60 copies of mitochondrial genomes were transmitted per *C. elegans* hermaphrodite per generation (table 4).

Discussion

To date, direct estimates of the overall spontaneous mtDNA mutation rate (μ_{total}) have only been obtained for five species including one unicellular and four multicellular eukaryotes (Denver et al. 2000; Haag-Liautard et al. 2008; Molnar et al. 2011; Saxer et al. 2012; Xu et al. 2012). μ_{total} ranges approximately 24-fold from 0.0676 to 1.63×10^{-7} changes per site per generation, with the protist *D. discoideum* and the microcrustacean *D. pulex* representing the lower and higher ends of this range, respectively. Our μ_{total} estimate of 1.05×10^{-7}/site/generation for *C. elegans* falls at the higher end of this range, but is slightly lower than that of *D. pulex* (1.63×10^{-7}/site/generation; Xu et al. 2012) and a preceding estimate for *C. elegans* that employed a direct sequencing approach spanning ~75% of the mtDNA genome (1.6×10^{-7}/site/generation; Denver et al. 2000). All of these mutation rate estimates have large confidence intervals and our overall estimate for the *C. elegans* mtDNA genome is not significantly different from that of Denver et al. (2000).

The spontaneous mtDNA base-substitution rate (μ_{bs}) and the spectrum of accrued molecular changes have been generated in seven species comprising two and five unicellular and multicellular eukaryotes, respectively. Estimates of μ_{bs} for the unicellular eukaryotes *S. cerevisiae* and the protist *P. tetraurelia* differ ~six-fold (1.22 vs. 6.96×10^{-8} base-substitutions per nucleotide site per generation, respectively) (Lynch et al. 2008; Sung et al. 2012). For the five multicellular eukaryotic species (*D. melanogaster*, *D. pulex*, and the nematodes *C. elegans*, *C. briggsae*, and *P. pacificus*), the spontaneous mtDNA base-substitution rate varies less than three-fold with a range of $3.75–9.7 \times 10^{-8}$ base-substitutions per nucleotide site per generation (Denver et al. 2000; Haag-Liautard et al. 2008; Howe et al. 2010; Molnar et al. 2011; Xu et al. 2012), with *Daphnia* exhibiting the lowest rate and a preceding estimate in *C. elegans* being the highest. Our direct estimate of the *C. elegans* mtDNA base-substitution rate $\mu_{bs} = 4.32 \times 10^{-8}$ per nucleotide site per generation is at the lower end of this range while extremely similar to that of the nematode *P. pacificus* (4.5×10^{-8} per site per generation) (Molnar et al. 2011). We note the existence of a greater than two-fold difference in the mtDNA μ_{bs} estimates from our analysis of *C. elegans* MA lines relative to that of Denver et al. (2000). Both studies utilized spontaneous MA lines of *C. elegans* derived from a single hermaphrodite of the laboratory *N2* Bristol stock. However, there are differences in the methodology that should be mentioned. Our sample set of 20 *C. elegans* MA lines was significantly smaller than the 74 MA lines analyzed by Denver et al. (2000). However, Denver et al. (2000) used a direct Sanger sequencing approach that spanned ~75% of the mtDNA genome, with MA lines that had been subjected to approximately 215 consecutive generations of MA. Our approach utilized the Illumina paired-end sequencing platform that spanned a near-complete mtDNA genome (~97%) in an MA experiment conducted over 409 consecutive generations. Because we were able to detect mutations present at

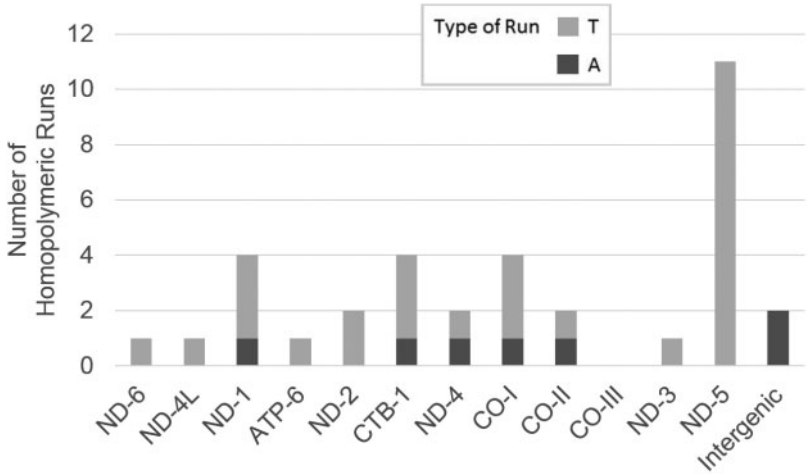

FIG. 7. The distribution of >6 bp A or T homopolymeric runs in protein-coding genes of the *C. elegans* mtDNA genome. The ND5 locus contains proportionally more homopolymeric runs than any other protein-coding gene within the genome.

lower frequencies than Denver et al. (2000) due to greater depth of sequencing, the discrepancy between our mtDNA mutation rate estimates is even larger than it appears.

The spontaneous mutation rate for indel events (μ_{indel}) in the *C. elegans* mtDNA genome is 6.14×10^{-8} changes/site/generation and is approximately 1.4 times higher than the spontaneous base-substitution rate. The ratio of indels to single-base substitutions was 1.7. A higher rate of indel events in the mtDNA genome was also observed for *D. pulex* (Xu et al. 2012) and the congeneric species *C. briggsae* (Howe et al. 2010). However, this pattern is not a generic feature of metazoan mtDNA genomes *per se*; notably *D. melanogaster* (Haag-Liautard et al. 2008) and the nematode *P. pacificus* (Molnar et al. 2011) exhibit higher rates of base-substitution relative to indel events.

In all of the major eukaryotic groups, although there exists a wide spread in the base composition of mitochondrial genomes, the majority of species exhibit a skew towards higher A + T content. All eight eukaryotic species, in which the spontaneous mtDNA process has been studied via MA lines, possess mtDNA genomes with highly biased A + T-rich base compositions ranging from 62% to 84%. Are these skewed compositions the result of a strongly biased mutation pressure towards A/T substitutions or do they also reflect countering forces (such as selection) to maintain an optimum equilibrium nucleotide composition? Insights into these questions can be gained by determining the spontaneous base mutation pressure in these genomes and to what extent the extant base composition deviates from the pattern of mutational input. The simplest explanation for the variable base composition of mtDNA genomes between species is that they arise from different biases in mutation rates. This hypothesis predicts that A + T-rich mitochondrial genomes should have higher rates of G/C→A/T mutations relative to A/T→G/C ones. However, a previous analysis of mutations in the *C. elegans* mitochondrial genome suggested a mutational bias towards a high G + C content and it was proposed that the high A + T content of the genome (76%; Okimoto et al. 1992) was due to natural selection (Denver et al. 2000). This G/C mutation bias in *C. elegans* reported by Denver et al. (2000) is at odds with the A/T bias observed in the spontaneous mtDNA mutation patterns reported in other multicellular eukaryotes studied to date; 57, 86, 87% spontaneous base substitutions increased A/T content in *P. pacificus* (Molnar et al. 2011), *D. melanogaster* (Haag-Liautard et al. 2008), *C. briggsae* (Howe et al. 2010), respectively. These conflicting patterns in the spontaneous base substitution spectrum of mtDNA were referred to as a "muddle of mutation across taxa" (Montooth and Rand 2008). The results of our study on the spontaneous mtDNA base substitution spectrum in *C. elegans* are incongruent with those of Denver et al. (2000). We demonstrate a strong G/C→A/T mutation bias for mtDNA base changes in *C. elegans*, which is corroborated with our analysis of the spectrum of nucleotide substitutions in 38 natural isolates of *C. elegans*. The equilibrium G + C content predicted from either our MA results or analysis of base substitutions in natural isolates is similar to the observed G + C content in the third codon position of protein-coding genes. Therefore, we do not find evidence to support the conclusion that the nucleotide composition at synonymous sites is under strong selection as previously suggested (Denver et al. 2000). In this respect, the base substitution mutation spectrum in the *C. elegans* mitochondrial genome is similar to its congeneric *C. briggsae* (Howe et al. 2010) as well as other multicellular eukaryotes studied to date (Haag-Liautard et al. 2008; Molnar et al. 2011).

The intracellular dynamics of mitochondrial mutations in *C. elegans* are more similar to that of other nematodes than previously appreciated. Previous work has sought to contrast *C. elegans* and *C. briggsae* with regards to the level of heteroplasmy of recent mitochondrial mutations, the conclusion being that heteroplasmy was common in *C. briggsae* while extremely rare in *C. elegans* (Howe et al. 2010). However, we find that heteroplasmic mutations are pervasive in *C. elegans*, as is the case in *Drosophila* (Haag-Liautard et al. 2008). Of a total of 46 new mtDNA variants identified across the three population size treatments in our MA experiment, only one variant in line 10E showed a frequency of 1 (tables 1–3). In fact, homoplasmic mutations were surprisingly rare given that the duration of the MA experiment reported here comprised almost two times the number of generations of previous experiments, thereby increasing the evolutionary time for novel mitochondrial mutations to reach fixation. In direct contrast, Denver et al. (2000) reported only one heteroplasmic mutation among the 26 spontaneous mtDNA variants identified in their set of *C. elegans* MA lines maintained at $N = 1$. The divergent conclusions regarding the extent of heteroplasmy in the *C. elegans* mtDNA genome between our study and that of Denver et al. (2000) may be attributed to technical challenges in ascertaining heteroplasmic variants, especially low-frequency ones, using traditional Sanger sequencing relative to next-generation sequencing methods. Based on the frequency distribution of mtDNA variants in our $N = 1$ MA lines, the effective number of mitochondrial genomes ($N_{e[mtDNA]}$) transmitted from the mothers to their offspring is ~60. Our $N_{e[mtDNA]}$ estimate for *C. elegans* may be conservative because our filtering screen excluded all variants in less than 3% frequency, some fraction of which may have been valid, rare heteroplasmies. However, our $N_{e[mtDNA]}$ estimate is larger than that for other model organisms such as fly (13–42 copies; Haag-Liautard et al. 2008), and *Daphnia* (5–10 copies; Xu et al. 2012) and brings into question a preceding conclusion that *C. elegans* experiences more severe mtDNA bottlenecks relative to *C. briggsae* (Howe et al. 2010). Although Howe et al. (2010) did not provide an estimate for the $N_{e[mtDNA]}$ in *C. briggsae*, the higher overall frequency of heteroplasmic variants and the common occurrence of low-frequency heteroplasmies in our study relative to their findings in *C. briggsae* would, at face value, argue for a wider bottleneck in *C. elegans*. One of the consequences of small mitochondrial bottleneck and low $N_{e[mtDNA]}$ is to increase the efficiency of selection against deleterious and selfish mutations (Bergstrom and Pritchard 1998; Rand 2011; Cooper et al. 2015). This raises the question of whether purifying selection against mitochondrial mutations operates less efficiently in organisms such as *C. elegans* compared with species with lower $N_{e[mtDNA]}$.

The estimates of mutation rates in MA experiments rest on the assumption that most mutations are effectively neutral. Given the complex population dynamics of mitochondria, where they exist in nested population hierarchies (multiple genomes within mitochondria, multiple mitochondria within a cell, multiple oocytes per female, multiple individuals in a population, and so forth), there is a potential for selection both between individuals and among mitochondria within individuals (Rand 2001). On the one hand, selection can operate within an individual to eliminate deleterious mutations. Therefore, many deleterious mitochondrial mutations may escape detection in MA experiments. On the other hand, many deleterious mutations may, in fact, have a competitive advantage within individuals (Ma and O'Farrell 2016). For example, deletions in *C. elegans* (Gitschlag et al. 2016) as well as ND5 deletions in *C. briggsae* (Clark et al. 2012) exhibit selfish drive. In this light, the high incidence of mutations in ND5 raises questions regarding the selective neutrality of these mutations in general. For example, in the $N = 1$ lines, approximately half of all mutations were detected in ND5 and all but one of these were either frameshift or nonsynonymous mutations. Overall, synonymous mutations were rare in our study and it remains a possibility that many frameshift and nonsynonymous mutations in high frequency were overrepresented due to selfish drive.

Our spontaneous MA experiment deviated in some aspects of its design from preceding MA experiments in other model organisms. In addition to maintaining *C. elegans* MA lines at the minimum bottleneck size possible ($N = 1$) for determination of the spontaneous rate of mutational input with little influence of selection, we additionally maintained MA lines at two larger population size treatments of $N = 10$ and 100 individuals ($n = 10$ and 5 lines, respectively). By determining the frequency distributions and patterns of mtDNA mutation in these two larger population size treatments as well as in 38 natural isolates, we aimed to investigate the influence of increasing intensities of natural selection on dictating the evolutionary dynamics and persistence of newly occurring mtDNA mutations. This analysis does come with some limitations. Our sample sizes for these larger treatments was limited (ten $N = 10$ and five $N = 100$ lines) due to logistic reasons associated with manually transferring 620 worms each generation during the course of the MA experiment. Furthermore, despite the order of magnitude higher mutation rate expected in metazoan mtDNA, the target size for mutations is relatively small in the mtDNA genome relative to the nuclear genome given the former's extremely compact size. The number of mitochondrial mutations detected in MA studies has therefore been small, ranging from 12 in *Pristionchus* (Molnar et al. 2011) to 36 in *D. melanogaster* (Haag-Liautard et al. 2008). The total number of mutations detected in our experiments was 46; 24 of these were found in the $N = 1$ lines. The challenges in detecting selection on spontaneous mitochondrial mutations under these conditions are a combination of (i) a small number of potentially deleterious mutations, and (ii) the possibility that mitochondrial mutations are likely to be recessive. It is thought that haplo-insufficient mitochondria need to reach a relatively high frequency (a critical threshold), even as high as 80%, for a deleterious phenotype to become apparent (Stewart and Chinnery 2015). Nonetheless, there is evidence that a fraction of the spontaneous nonsynonymous and frameshift mutations are indeed deleterious and selected against in the larger population lines ($N = 10$ or 100 individuals). There is a significant negative correlation between the frequency of spontaneous nonsynonymous/frameshift mutations and population size, which suggests that selection keeps some intracellular variants at low frequencies in the larger-sized experimental populations ($N = 10$ and 100 individuals). Our analysis of the published mtDNA genomes of the 38 natural isolates was conducted to investigate the distribution of mtDNA variants in populations presumably facing the most stringent level of natural selection. Natural isolate mtDNA genomes possess the lowest proportion of nonsynonymous mutations among the various lines examined in this study. This observation is in accord with the expectation that nonsynonymous mutations have deleterious fitness consequences and are subject to strong purifying selection. The $N = 1$ MA lines possess a greater fraction of mtDNA variants in intermediate frequencies than present in natural isolates, the latter showing a preponderance of extremely low- or high-frequency variants. This could be a combination of two factors: (i) natural selection suppressing the intracellular frequencies of deleterious variants in the wild, and/or (ii) insufficient evolutionary time for new mtDNA variants in the $N = 1$ MA lines to reach fixation via genetic drift.

In conclusion, the combination of massively parallel sequencing technology and MA experiments is a powerful approach to address long-standing questions about the frequency and spectrum of spontaneous mutations. In the case of mtDNA genomes, next-generation sequencing has proved to be particularly useful for detecting low-frequency heteroplasmic mutations. Most of the mutations detected in this study had not yet reached fixation within cells to become homoplasmic during the course of the experiment. However, there was evidence that selection is operating to reduce the intracellular frequency of potentially deleterious nonsynonymous and frameshift mutations in populations of limited size ($N = 10$ and 100 individuals). Furthermore, by comparing the results from traditional MA experiments (or MA populations with differing N_e as was done in this study) to standing genetic variation in natural populations, the dual role of natural selection and mutation pressure operating to shape genetic variation can be investigated. Understanding the population dynamics of heteroplasmic variants is important both for population and evolutionary genetics as well as for the study of pathogenic mitochondrial mutations that contribute to human diseases.

Materials and Methods

Mutation Accumulation Experiment

MA lines of hermaphroditic *Caenorhabditis elegans* were propagated at three different constant population sizes ($N = 1$, 10, and 100) for 409 generations, as described in Katju et al. (2015). Twenty lines were propagated through

single hermaphrodite ancestry each generation where possible (lines 1A–1T), while 10 (lines 10A–10J) and five lines (100A–100E) were passaged through bottlenecks of 10 and 100 hermaphrodites each generation, respectively. Each new MA generation was established by randomly selecting the designated number of founder hermaphrodites as L4 larvae and manually transferring to a new plate. Worm populations were cultured on Nematode Growth Medium at 20 °C on either 60 × 15 mm (lines maintained at $N = 1$ and 10) or 90 × 15 mm Petri dishes (for populations of size $N = 100$) seeded with 250 μL or 750 μL of *Escherichia coli* OP50 in YT medium.

Stocks of the MA lines were cryogenically preserved throughout the course of the experiment, at approximately every 50 MA generations. Following the termination of the mutation accumulation experiment at MA generation 409, all extant lines were cryogenically preserved at −86°C. Three $N = 1$ lines had gone extinct prior to the termination of the experiment (line 1H at MA generation 293; 1S at MA generation 328, and line 1T at MA generation 309) but pre-extinction frozen stocks were available for the purpose of genome sequencing.

DNA Extraction and Sequencing

The ancestral pre-MA control and 35 MA lines were thawed at room temperature on standard NGM plates. One, four and five worms were randomly selected and sequestered individually on to fresh plates from the $N = 1$ and control, $N = 10$, and $N = 100$ lines, respectively. A single hermaphrodite per $N = 1$ line was selected for sequencing, yielding 20 $N = 1$ genomes. For the three $N = 1$ lines that became extinct during the course of the MA experiment, we thawed worms from the latest frozen stock available (line 1H at 281 MA generations, line 1S at 291 MA generations, and line 1T at 305 MA generations). MA lines maintained at larger population sizes were expected to harbor greater standing genetic variation. Hence, we initiated sequencing the genomes of four randomly picked individuals from each $N = 10$ line (total 10 lines), in order to generate 40 (4 × 10) genomes. Similarly, five individuals from each $N = 100$ line (total 5 lines) were selected for generating 25 (5 × 5) genomes. A single genome of the ancestral pre-MA control was submitted for whole-genome sequencing (WGS henceforth). This design yielded 86 nematode lines for WGS (fig. 1). The 86 thawed, isolated worms were each transferred to an agarose plate with a HB101 bacterial lawn. A population of >5000 individuals all descended from the initial single worm was allowed to grow on the plate till starved, and the tissue collected for gDNA extraction.

Genomic DNA was isolated using a previously described protocol (Flibotte et al. 2010; Thompson et al. 2013) with a PureGene Genomic DNA Tissue Kit (QIAGEN no. 158622) following a supplemental nematode protocol. DNA quality was tested by electrophoresis on 1% agarose gels and DNA concentration was determined both by Thermo Fisher Nanodrop spectrophotometer and by BR Qubit assay (Invitrogen). Two micrograms of each sample was sonicated in 85μL TE buffer to yield 200–400 bp fragments (Covaris, E-Series), and then end-repaired using the NEBNext end repair module (New England BioLabs) and purified with Agencourt AMPure XP beads (Beckman Coulter Genomics). Beads were left in the reaction until adapter ligation as previously described (Thompson et al. 2013). Subsequent steps added 3′ adenine overhangs (AmpliTaq DNA Polymerase Kit, Life Technologies) and ligated custom pre-annealed Illumina adapters. Samples were then PCR amplified with Kapa Hifi DNA Polymerase (Kapa Biosystems) using Illumina's paired-end genomic DNA primer with an 8 bp barcode. PCR products were size-fractionated on 6% PAGE gels and 300–400 bp fractions were excised, gel extracted by diffusion at 65 °C, and gel filtered (NanoSep, Pall Life Sciences) before final purification with Agencourt AMPure beads. DNA quality and quantification were assessed using Agilent HS Bioanalyzer and HS Qubit assays. Multiplexed DNA libraries were sequenced via Illumina HiSeq at the Northwest Genomics Center at the University of Washington, using default quality filters for sequence reads. All reads generated in this project have been deposited on the NCBI Sequence Read Archive under Project ID PRJNA352429 (further details and accession numbers are available in supplementary table S4, Supplementary Material online).

Sequence Alignment and Variant Detection

The reads of the 86 genomes sequenced above were demultiplexed using a custom Perl script, and raw reads for each strain were stored in FastQ files. The reads for each MA individual sequenced and every natural isolate were individually aligned to the reference N2 genome (version WS247; www.wormbase.org) using the Burrows-Wheeler Aligner (BWA Version 0.5.9) (Li and Durbin 2009) and Phaster (Philip Green, personal communication), with default settings. Resulting alignments were sorted via SAMtools (Version 0.1.18; Li and Durbin 2009), and the rmdup utility of the same package was used in order to remove potential PCR and optical duplicates. Potential variants were identified using the mpileup utility of SAMtools, while filtering of the variants was subsequently performed using a custom script (see below for details). Additional potential variants were identified using the methods previously described (Rebolledo-Jaramillo et al. 2014). All potential variants with base qualities less than 30 were removed from the dataset. Variants called at sites with a read depth of less than 10% of the average read depth at the same site across all strains were discarded. We additionally discarded variants meeting the following criteria: (i) pre-existing variants in the ancestral control line, (ii) SNPs below a threshold of 3% of the read depth, or those which were supported by less than three reads. Every variant was verified manually using the Integrative Genomics Viewer (Robinson et al. 2011; Thorvaldsdóttir et al. 2013) before further analysis.

Variants reaching a threshold of 80% of all reads in at least one individual within a given population were considered to be polymorphic, while those occurring below this threshold are referred to as heteroplasmic. All accepted variants in the MA lines and natural isolates were annotated using the Wormbase gene annotations (version WS247) (Harris et al. 2010) and classified as protein-coding, noncoding RNA

(rRNA or tRNA), untranslated regions (intergenic regions or 3′-/5′-untranslated regions), or intergenic regions.

Independent Validation of mtDNA Variants in $N = 1$ Lines

The putative list of 24 mtDNA variants identified from the WGS data for the $N = 1$ MA lines (table 1) were further validated using one of three approaches. First, the two large deletions of 107 and 499 bp in lines 1C and 1G, respectively, were directly confirmed by PCR and gel electrophoresis. Second, 16 SNPs and small indel variants were genotyped using the KASPar allele-specific fluorescent genotyping system (KBioscience, www.kbioscience.co.uk) as previously described (Seabury et al. 2010; Smith and Maughan 2015). The KASPar genotyping system employs fluorescence to measure abundance of allele-specific extension products. SNP-specific assays were designed by KBioscience and primer sequences are available on request. Genotype clustering and calling was performed using KlusterCaller software (Kbiosciences) and the endpoint genotyping module incorporated within the Roche LC480 (Roche Applied Science). Third, three additional SNPs/small indels were verified via droplet digital PCR (ddPCR) as previously described (Hindson et al. 2011; Rebolledo-Jaramillo et al. 2014). Three SNPs from the $N = 1$ lines were not verified by these methods because of challenges associated with designing appropriate primers. However, in light of our success in verifying every other variant for which primers could be constructed, we treated these as real mutations. Mutations in the $N = 10$ and $N = 100$ were all found in multiple individuals within the same populations and were therefore not subjected to additional verification.

Analysis of Mitochondrial Variants in Natural Isolates of *C. elegans*

The raw reads for mitochondrial genomes of 38 previously sequenced *C. elegans* natural isolates (supplementary table S2, Supplementary Material online) (Thompson et al. 2013) were downloaded from the Sequence Read Archive (Leinonen et al. 2011). These genomes were aligned against the mitochondrial genomes of *C. briggsae*, *C. tropicalis* and *C. nigoni* (strain JU1421) using MUSCLE (Edgar 2004). Subsequently, a maximum likelihood tree was reconstructed in MEGA with default settings, which was rooted with *C. briggsae*, *C. tropicalis*, and *C. nigoni* as outgroups (Tamura et al. 2013) (supplementary fig. S1, Supplementary Material online). Ancestral state reconstruction for each position was performed in R (R Core Team 2014) using the *ace* function in the APE package (Paradis et al. 2004) allowing for different rates between substitutions. The directionality of a given mutation was determined by identifying the nucleotide with the highest likelihood at the node of the last common ancestor of the shared mutation. Polymorphic mutations present in multiple descendants of their last common ancestor were treated as a single mutation. Sites that differed between the two major clades in the phylogenetic tree of the *C. elegans* mitochondrial genome were not used to assign directionality of nucleotide substitutions, although they were used to calculate transition/transversion ratios.

Calculation of Mutation Rates and the Mitochondrial Effective Population Size ($N_{e[mtDNA]}$)

In order to calculate the mutation rates for the mutation accumulation lines, we assumed that all variants detected are under mutation-drift balance, in which case the probability of any mutation reaching fixation is equal to its observed frequency in the population within which it occurs (Kimura 1983; Haag-Liautard et al. 2008; Saxer et al. 2012). Hence, the line-specific mutation rates were calculated in such a way that each variant (both heteroplasmic and polymorphic) contributes to the total mutation rate in proportion to the frequency at which it was observed in the dataset:

$$\mu_i = \frac{\sum \frac{f(m_j)}{n_i}}{b_i * g_i}$$

where μ_i represents the mutation rate for each individual population, $f(m_j)$ equals the frequency of each individual variant within the population, n_i is the number of individuals sequenced for the population (1, 4, or 5), and b_i and g_i refer to the number of bases analyzed and the number of generations for the population, respectively. The final spontaneous mutation rate was calculated by averaging the individual mutation rates for lines of population size $N = 1$. 95% confidence intervals for the spontaneous mutation rates were determined via bootstrapping. Line-specific mutation rates were randomly sampled with replacement to recalculate the mutation rate for 20 strains for a total of 10,000 iterations.

We also employed a maximum likelihood (ML) model (Haag-Liautard et al. 2008) to infer the effective population size of mitochondria transmitted to the progeny each generation. The mitochondrial effective population size, $N_{e[mtDNA]}$, was modeled using a haploid Wright–Fisher transition matrix method which is discussed in detail in Haag-Liautard et al. (2008).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Author Contributions

V.K. conceived and designed the experiments; V.K. created the experimental evolution lines; V.K. and U.B. jointly supervised the study; O.T, R.H.W., and D.G.M. generated the WGS data; A.K. filtered and collected the data; A.K., U.B., P.D.K., V.K. analyzed the data; P.D.K. contributed to ML analyses; A.K., U.B., and V.K. wrote the article; all authors revised and approved the article.

Acknowledgments

The research and publication costs were supported by a National Science Foundation grant NSF-MCB 1330245 from the Division of Molecular and Cellular Biosciences to VK. We owe a special thanks to Philip Green from the University of Washington for providing the program Phaster. We thank Lucille Packard for her assistance in the creation of the mutation accumulation lines, Drew Hillhouse and Chris Seabury

for help with variant validation, and Boris Rebolledo-Jaramillo for providing a list of variants identified using an alternative approach. The wild-type N2 nematode strain used in this study was provided by the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health (NIH) National Center for Research Resources (NCRR).

References

Aanen DK, Spelbrink JN, Beekman M. 2014. What cost mitochondria? The maintenance of functional mitochondrial DNA within and across generations. *Philos Trans R Soc Lond B Biol Sci*. 369:20130438.

Avise JC. 1994. Molecular markers, natural history, and evolution. New York: Chapman & Hall.

Avise JC. 2000. Phylogeography: the history and formation of species. Cambridge: Harvard University Press.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst*. 18:489–522.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol*. 13:729–744.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol*. l68:39–50.

Bergstrom CT, Pritchard J. 1998. Germline bottlenecks and the evolutionary maintenance of mitochondrial genome. *Genetics* 149:2135–2146.

Birky CW.Jr 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms and models. *Annu Rev Genet*. 35:125–148.

Bowe LM, Coat G, dePamphilis CW. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc Natl Acad Sci U S A*. 97:4092–4097.

Brown WM, George M Jr, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A*. 76:1967–1971.

Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.

Chaw S-M, Chang C-C, Chen H-L, Li W-H. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol*. 58:424–441.

Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc Natl Acad Sci U S A*. 101:17741–17746.

Clark KA, Howe DK, Gafner K, Kusuma D, Ping S, Estes S, Denver DR. 2012. Selfish little circles: transmission bias and evolution of large deletion-bearing mitochondrial DNA in *Caenorhabditis briggsae* nematodes. *PLoS One* 7:e41433.

Cooper BS, Burrus CR, Ji C, Hahn MW, Montooth KL. 2015. Similar efficacies of selection shape mitochondrial and nuclear genes in both *Drosophila melanogaster* and *Homo sapiens*. *G3 Genes|Genomes|Genetics* 5:2165–2176.

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.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

Eyre-Walker A, Smith NH, Smith JM. 1999. How clonal are human mitochondria? *Proc Biol Sci*. 266:477–483.

Flibotte S, Edgley ML, Chaudhry I, Taylor J, Neil SE, Rogula A, Zapf R, Hirst M, Butterfield Y, Jones SJ, et al. 2010. Whole-genome profiling of mutagenesis in *Caenorhabditis elegans*. *Genetics* 185:431–441.

Gaut BS. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *Evol Biol*. 30:93–120.

Gitschlag BL, Kirby CS, Samuels DC, Gangula RD, Mallal SA, Patel MR. 2016. Homeostatic responses regulate selfish mitochondrial genome dynamics in *C. elegans*. *Cell Metab*. 24:91–103.

Gyllensten U, Wharton D, Josefsson A, Wilson AC. 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature* 352:255–257.

Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol*. 6:e204.

Hall HG, Muralidharan K. 1989. Evidence from mitochondrial DNA that African honey bees spread as continuous maternal lineages. *Nature* 339:211–213.

Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu Rev Ecol Evol Syst*. 40:151–172.

Harris TW, Antoshechkin I, Bieri T, Blasiar D, Chan J, Chen WJ, De La Cruz N, Davis P, Duesbury M, Fang R, et al. 2010. WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res*. 38:D463–D467.

Havird JC, Sloan DB. 2016. The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Mol Biol Evol*. 33:3042–3053.

Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, et al. 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem*. 83:8604–8610.

Howe DK, Baer CF, Denver DR. 2010. High rate of large deletions in *Caenorhabditis briggsae* mitochondrial genome mutation processes. *Genome Biol Evol*. 2:29–38.

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

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kondo R, Satta Y, Matsuura ET, Ishiwas H, Takahata N, Chigusa SI. 1990. Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics* 126:657–663.

Kvist L, Martens J, Nazarenko AA, Orell M. 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Mol Biol Evol*. 20:243–247.

Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. 2011. The sequence read archive. *Nucleic Acids Res*. 39:D19–D21.

Lemire B. 2005. Mitochondrial genetics. *Wormbook* 14:1–10.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Lynch M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol Biol Evol*. 13:209–220.

Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* 311:1727–1730.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A*. 105:9272–9277.

Ma H, O'Farrell PH. 2016. Selfish drive can trump function when animal mitochondrial genomes compete. *Nat Genet*. 48:798–802.

McCauley DE, Bailey MF, Sherman NA, Darnell MZ. 2005. Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* 95:50–58.

Molnar RI, Bartelmes G, Dinkelacker I, Witte H, Sommer RJ. 2011. Mutation rates and intraspecific divergence of the mitochondrial genome of *Pristionchus pacificus*. *Mol Biol Evol*. 28:2317–2326.

Montooth KL, Rand DM. 2008. The spectrum of mitochondrial mutations differs across species. *PLoS Biol*. 6:e213.

Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst*. 18:269–292.

Neale DB, Marshall KA, Sederoff RR. 1989. Choloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don. Endl. *Proc Natl Acad Sci U S A*. 86:9347–9349.

Neiman M, Taylor DR. 2009. The causes of mutation accumulation in mitochondrial genomes. *Proc Biol Sci*. 276:1201–1209.
Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130:471–498.
Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol*. 28:87–97.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Rand DM. 2001. The units of selection on mitochondrial DNA. *Annu Rev Ecol Syst*. 32:415–448.
Rand DM. 2008. Mitigating mutational meltdown in mammalian mitochondria. *PLoS Biol*. 6:e35.
Rand DM. 2011. Population genetics of the cytoplasm and the units of selection on mitochondrial DNA in *Drosophila melanogaster*. *Genetica* 139:685–697.
Rebolledo-Jaramillo B, Su MS, Stoler N, McElhoe JA, Dickins B, Blankenberg D, Korneliussen TS, Chiaromonte F, Nielsen R, Holland MM, et al. 2014. Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A*. 111:15474–15479.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol*. 29:24–26.
Sajantilla A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S. 1996. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc Natl Acad Sci U S A*. 93:12035–12039.
Saxer G, Havlak P, Fox SA, Quance MA, Gupta S, Fofanov Y, Strassmann JE, Queller DC. 2012. Whole genome sequencing of mutation accumulation lines reveals a low mutation rate in the social amoeba *Dictyostelium discoideum*. *PLoS One* 7:e46759.
Seabury CM, Seabury PM, Decker JE, Schnabel RD, Taylor JF, Womack JE. 2010. Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. *Proc Natl Acad Sci U S A*. 107:151–156.
Smith SM, Maughan PJ. 2015. SNP genotyping using KASPar assays. *Methods Mol Biol*. 1245:243–256.
Stewart JB, Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet*. 16:530–542.
Sueoka N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc Natl Acad Sci U S A*. 48:582–592.
Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, Lynch M. 2012. Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. *Proc Natl Acad Sci U S A*. 109:19339–19344.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.
Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, Au V, Chaudhury I, Fernando L, Hutter H, et al. 2013. The million mutation project: a new approach to genetics in *Caenorhabditis elegans*. *Genome Res*. 23:1749–1762.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 14:178–192.
Wallace DC, Chalkia D. 2013. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Biol*. 5:a021220.
White DJ, Wolff JN, Pierson M, Gemmell NJ. 2008. Revealing the hidden complexities of mtDNA inheritance. *Mol Ecol*. 17:4925–4942.
Wolfe KH, Li W-H, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proc Natl Acad Sci U S A*. 84:9054–9058.
Xu S, Schaack S, Seyfert A, Choi E, Lynch M, Cristescu ME. 2012. High mutation rates in the mitochondrial genomes of *Daphnia pulex*. *Mol Biol Evol*. 29:763–769.