**SUPPLEMENTARY MATERIAL**

**A picture containing text, opener, tool

Description automatically generated**

**Figure S1.** Culturing design used to obtain the nine replicates assayed in this study.

****

**Figure S2.** The results of Figure 1a (concerning the mutation spectrum) remain unchanged when excluding the 18 variants (at 15 sites) that are shared among replicates from the population. The excluded variants are listed in Table S3. For all excluded variants, the coverage of the reference base was also adjusted so that reference base coverage at excluded sites had no contribution to substitution class frequency. Removing shared sites has no impact the observation that CG🡪TA transitions and CG🡪AT transversions dominate the mutation spectrum. In addition, we still find significant variation in the (log transformed) SNV frequency between the six substitution classes (one-way ANOVA, *p* << 0.0001).

****

**Figure S3. Average DCS depth as a function of GC content.** Each point on the plot depicts a 50-bp window, with DCS depths averaged across the nine replicates and error bars reporting one standard error. Points were jittered for clarity.

****

**Figure S4.** Variation in total SNV frequencies across the 12 protein-coding genes for all six substitution classes (one way ANOVA, *p* = 0.0072). In separate tests with each substitution class, significant between gene variation was observed only for only CG🡪AT transversions and CG🡪TA transitions (see Figure 2c).

****

**Figure S5.** Correlation between gene specific SNV frequencies and GC content for CG🡪AT transversions andCG🡪TA transitions. R and *p* values (shown in legend) are from a Pearson correlation, implemented in R with the cor.test command.

Chart, scatter chart

Description automatically generated

**Figure S6.** C🡪T transition and G🡪T transversion enrichment values across the length of the *C. elegans* mtDNA. C🡪T enrichment is plotted in dark blue and was calculated as:

(C🡪T frequency - G🡪A frequency) / (C🡪T frequency + G🡪A frequency).

G🡪T enrichment is plotted in tan and was calculated as:

(G🡪T frequency - C🡪A frequency) / (G🡪T frequency + C🡪A frequency).

Enrichment values were calculated for 1000 bp windows (step size 100 bp), with each point on the plot showing the enrichment value for a given window. The point size is scaled to the total number of mutations observed in a window, according to the key in the bottom left of the figure. Windows with 0 observed mutations (undefined enrichment values) are omitted from the plot. The regression lines were generated with the geom\_smooth function in ggplot2 using local fitting (i.e. method = loess). In the schematic gene map at the bottom of the x-axis (generated with OGDRAW; Greiner *et al.* 2019), protein coding genes are shown in tan, tRNAs are shown in blue, rRNAs are shown in red and the intergenic regions are shown in pink.



**Figure S7.** As in Figure 5, except that shared indels are assumed to be shared due to common ancestry, so are only counted once regardless of the number of replicates in which they occur.

**Table S1. DCS coverage and percent mapping for each replicate.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Replicate | Total DCS bp | Percent reads Mapped | Coverage per mtDNA bp | SRA Accession numbers |
| 1a | 245145853 | 99.95 | 15451 | SRR14352243 |
| 1b | 143958801 | 99.97 | 9099 | SRR14352244 |
| 1c | 242990460 | 99.95 | 15351 | SRR14352245 |
| 2a | 168251787 | 99.86 | 10691 | SRR14352246 |
| 2b | 203908065 | 99.93 | 12980 | SRR14352240 |
| 2c | 230152257 | 99.89 | 14624 | SRR14352241 |
| 3a | 236482816 | 99.86 | 14911 | SRR14352242 |
| 3b | 198009375 | 99.91 | 12462 | SRR14352247 |
| 3c | 218583804 | 99.94 | 13749 | SRR14352248 |
| Total | 1887483218 | 99.91 | 119319 | NA |

**Table S2. Positions which differ between the published N2 mitochondrial genome (NC\_001328.1) and our N2 line.** The two SNPs were completely fixed compared to NC\_001328.1. In contrast, the 10 bp indel was supported by the majority of all DCSs, but it was not completely fixed in any of the nine replicates.

|  |  |  |  |
| --- | --- | --- | --- |
| Type of difference | Position in NC\_001328.1 | Reference base in NC\_001328.1 | Reference base in our N2 line |
| SNP | 8429 | A | G |
| SNP | 12998 | C | T |
| indel | 3235 | A11 (11 bp homopolymer) | A10 (10 bp homopolymer) |

**Table S3. Variants excluded from Figure S2 that were shared in multiple replicates.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| replicate | mtDNA position | reference | variant | single stranded substitution type on F-strand | double stranded substitution type | reference base DCS count | variant DCS count |
| 3b | 1291 | C | T | C🡪T | CG🡪TA | 11431 | 1 |
| 2b | 1548 | A | G | A🡪G | AT🡪GC | 22415 | 1 |
| 3b | 1630 | G | T | G🡪T | CG🡪AT | 25562 | 1 |
| 3a | 1640 | G | T | G🡪T | CG🡪AT | 27766 | 1 |
| 3b | 1875 | G | A | G🡪A | CG🡪TA | 23486 | 1 |
| 3a | 5062 | C | T | C🡪T | CG🡪TA | 27428 | 1 |
| 1c | 5079 | G | A | G🡪A | CG🡪TA | 38890 | 158 |
| 2a | 5079 | G | T | G🡪T | CG🡪AT | 18278 | 1 |
| 2a | 5079 | G | A | G🡪A | CG🡪TA | 18278 | 5 |
| 3c | 5079 | G | A | G🡪A | CG🡪TA | 34029 | 4 |
| 3c | 5080 | C | T | C🡪T | CG🡪TA | 33887 | 1 |
| 2a | 5976 | C | T | C🡪T | CG🡪TA | 16256 | 1 |
| 3a | 8395 | C | T | C🡪T | CG🡪TA | 27410 | 1 |
| 3b | 8794 | T | C | T🡪C | AT🡪GC | 32385 | 1 |
| 1c | 8999 | T | C | T🡪C | AT🡪GC | 34443 | 1 |
| 2c | 8999 | T | C | T🡪C | AT🡪GC | 24974 | 1 |
| 2c | 9144 | C | T | C🡪T | CG🡪TA | 20784 | 1 |
| 2c | 12069 | G | A | G🡪A | CG🡪TA | 13551 | 1 |

**Table S4. Proportions of the counts of each substitution class of the total counts for all SNVs, singleton SNVs, and multi-DCS SNVs.** Note these proportions are not normalized for biased base composition of the *C. elegans* mtDNA or for differential probability of detection for AT and CG increasing variants due to uneven AT and CG coverage. For a normalized spectrum, see Figure 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Substitution Class | All SNVs | Singleton SNVs | Multi-DCS SNVs |
| AT🡪CG | 0.01 | 0.01 | 0.00 |
| AT🡪GC | 0.20 | 0.20 | 0.22 |
| AT🡪TA | 0.08 | 0.07 | 0.13 |
| CG🡪AT | 0.31 | 0.31 | 0.35 |
| CG🡪GC | 0.03 | 0.03 | 0.00 |
| CG🡪TA | 0.37 | 0.38 | 0.30 |
| Total Counts | 253 | 230 | 23 |

**Table S5. Proportions of the counts of each substitution class of the total counts for all SNVs as well as from the comparison of mtDNAs from 38 *C. elegans* natural isolates** (Thompson *et al.* 2013; Konrad *et al.* 2017)**.** Because substitutions in the population data cannot be reliably polarized, all substitution classes have been collapsed into four reversable classes (one transition and three transversions). As in Table S2, the proportions are not normalized to reflect the base composition of the mtDNA.

|  |  |  |  |
| --- | --- | --- | --- |
| Substitution Class | All SNVs | Natural isolates all substitutions | Natural isolates substitutions at four-fold degenerate sites |
| AT↔GC | 0.57 | 0.83 | 0.77 |
| AT↔CG | 0.32 | 0.07 | 0.09 |
| AT↔TA | 0.08 | 0.10 | 0.15 |
| GC↔CG | 0.03 | 0.002 | 0.00 |
| Total Counts | 253 | 408 | 162 |

**SUPPLEMENTARY REFERENCES:**

Greiner S., P. Lehwark, and R. Bock, 2019 OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47: 59–64.

Konrad A., O. Thompson, R. H. Waterston, D. G. Moerman, P. D. Keightley, *et al.*, 2017 Mitochondrial mutation rate, spectrum and heteroplasmy in Caenorhabditis elegans spontaneous mutation accumulation lines of differing population size. Mol. Biol. Evol. 34: 1319–1334.

Thompson O., M. Edgley, P. Strasbourger, S. Flibotte, B. Ewing, *et al.*, 2013 The million mutation project: A new approach to genetics in Caenorhabditis elegans. Genome Res. 23: 1749–1762.