Supplemental information on bioinformatic analysis

Image analysis, basecalling and error calibration was performed using standard Illumina software. Alignments to the annotated sequence of *C. elegans* available at WormBase WS225 were performed by BWA. Samtools was used for SNP and indel calling, with BAQ calculation turned off. All non-unique SNPs and indels are considered to be pre-existing and were filtered out using custom Perl scripts. We considered a SNP to be real if at least 80% of the called bases were non-wildtype for SNPs that are covered ≥4 times. As a second filter at least one of MA lines of the same genotype should be called as wildtype: having a coverage ≥4 of which ≥80% was of wildtype nature according to pileup generated with mpileup. Sanger sequencing of predicted SNPs validated these criteria.

To identify microsatellite mutations and deletions we used Pindel, developed by Ye et al[18]. Standard settings were used, except for changing the maximum allowed mismatch rate (-u) from 0.02 to 0.01. In addition, the sequencing error rate (-e) was adjusted to 0.0001. Pindel results were filtered using custom Java software that selects structural variations that are unique for one sample or the variation was supported by more than eight reads in one sample and eight times less often seen in any other sample. The latter criterium allowed us to also detect variations in homopolymers, which frequently contain sequencing errors. A deletion or deletion with template insertion of ≥20 nt was selected if the local average coverage was below five, while the directly surrounding area was covered >20 fold. Deletions and deletions with template insertions of <20 nt were included when both Pindel

and SAMtools reported the variation and SAMtools reported it to be homozygous (≥80% non-wildtype). A representative set of deletions was validated by Sanger sequencing. The complete sequencing data have been submitted to the NCBI Sequence Read Archive (SRA) with accession ID SRP020555.

The simulated set was created in two steps. First, the N2 reference genome was edited by creating random deletions with a size distribution similar to *polh-1;polk-1* throughout the genome using a custom Perl script. Second, the software tool wgsim was used to generate pair-end reads with a 100x coverage based on the modified genome. Subsequently, the reads were analyzed using the same pipeline as used for the MA lines.

For the analysis of the natural isolates paired-end whole genome sequence data was downloaded from the NCBI Sequence Read Archive (SRP011413) [38], and sequence reads were mapped to the *C. elegans* reference genome (Wormbase release 225). The average base coverage was 176x, 164x, 166x and 75x for AB2, CB4857, RC301 and CB4856, respectively. Pindel was used to detect structural variations (SVs) in the natural isolates as compared to the N2 reference genome. We included only SVs that had at least 10 reads supporting the SV and no reads supporting the reference genome. Next, we grouped all SVs from AB2, CB4857, RC301 together as the phylogenetic tree indicate that they are closely related (as compared to CB4856. As a second criterion, we collected only those SVs that were N2-like in CB4856, but were flagged as a SV in the AB2, CB4857 or RC301 group (and vice versa). We found 6,279 and 5,028 deletions of at least 1bp for CB4856 and the combined group that consists of CB4857, AB2 and RC301, respectively. The

regression curves displayed in figure 4 were created by iterating over the natural isolate datasets.

**Table S1. Whole genome sequencing statistics.**

| genotype | sample | # generations | # reads | average coverage | # bp >= 4x covered |
|---|---|---|---|---|---|
| N2 | N2 | 60 | 45,258,326 | 28x | 100,140,732 |
| | N4 | 60 | 23,693,826 | 16x | 99,675,920 |
| *polh-1(lf31)* | H7 | 60 | 46,203,688 | 39x | 100,229,062 |
| | H8 | 60 | 44,982,616 | 37x | 100,238,324 |
| *polk-1(lf29)* | K1 | 60 | 41,517,548 | 21x | 99,970,233 |
| | K4 | 60 | 39,275,458 | 30x | 100,235,635 |
| | K9 | 60 | 40,037,564 | 24x | 100,120,773 |
| *polh-1(lf31);polk-1(lf29)* | D4 | 32 | 46,284,780 | 21x | 99,911,564 |
| | D13 | 25 | 38,712,292 | 29x | 100,224,845 |
| | D14 | 25 | 59,163,976 | 27x | 100,202,641 |
| *msh-6 (pk2504)* | M13 | 10 | 48,338,722 | 19x | 99,236,278 |
| | M15 | 10 | 44,129,942 | 12x | 99,799,729 |

**Table S2. *unc-22* deletions in *polh-1polk-1* and *polh-1polk-1polq-1.***

| | size | left flank | deletion left | deletion right | right flank | insertion |
|---|---|---|---|---|---|---|
| *polh-1polk-1* | | | | | | |
| A | 83 bp | GTACCTACTCA | CGTCCAAATG | TTATCGAAAA | GAACGTGTGC | - |
| B | 74 bp | AATCCAGAAGT | CGATGACACC | CTTGGTTAGT | TATTTTTTGG | - |
| C | 153 bp | ACAAGGCTGGG | CCTGGACAAC | TAAAGGCTGG | AGCCACTGTT | - |
| D | 119 bp | GACTATCAAGG | CTGGTCAATC | TGATAACCCA | GAATACCAAT | AATCTGACTATCAAAGGAAATC TCAAGAATCTGACTATCAAAG |
| E | 93 bp | CTTGCAAAGGA | TCCATTTGGA | CACGTGACAA | CGGTGGATCA | - |
| F | 71 bp | TGTGAAGCCTT | ACGGAACTGA | ACCACCAGTT | GTTACTTGGC | - |
| G | | not identified | | | | |
| *polh-1polk-1 polq-1* | | | | | | |
| A | >4.7 kb | | | | | |
| B | >30.5 kb | | | | | |
| C | 19 - 20.6 kb | | | | | |
| D | 12660 bp | AAATGAGCACA | CTATTCTGTG | GAACAGGAGC | ATTTGGAGTT | |
| E | > 23.7 kb | | | | | |
| F | | not identified | | | | |

**Table S3. Frequency of *unc-22* mutations in *polq-1, polh-1polk-1* and *polh-1polk-1polq-1*.**

| Strain | | total # plates scored | # plates containing one or more twitchers | estimated mutation rate |
|---|---|---|---|---|
| N2 | wildtype | 40 | 0 | 0.00E+00 |
| XF152 | *polq-1* | 40 | 0 | 0.00E+00 |
| XF507 | *polh-1polk-1* | 46 | 7 | 8.00E-06 |
| XF840 | *polh-1polq-1polk-1* | 39 | 6 | 8.00E-06 |

**Table S4. Sequence analysis of reversion mutants for *unc-93(e1500)*.**

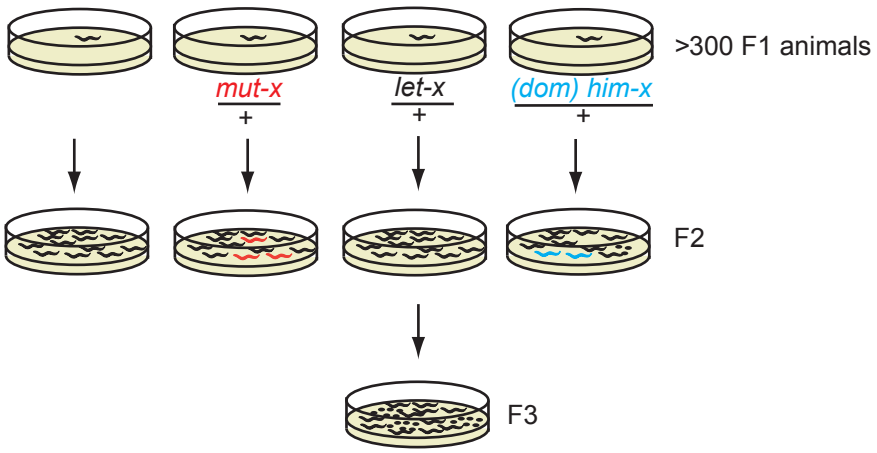
| | | | |
|---|---|---|---|
| wildtype | | | |
| *unc-93* | | | |
| deletions > 5kb | 0 | | |
| substitutions | 6 | cagttt(g>a)tctggc; C>Y | |
| | | gacacg(t>a)cacagt; V>D | |
| | | tgtctg(g>c)aatact; G>A | |
| | | aaatat(c>t)gatttt; R>L | |
| | | ggaatc(g>a)cggctt; T>A | |
| | | tgttag(g>t)taatgg; splice | |
| other | 1 | gaatat(tcga>deleted)aaaactt | 3bp > deletion > 12 bp |
| *sup-9* | | | |
| deletions > 5kb | 2 | | |
| substitutions | 1 | ccattg(g>a)gactta; G>stop | |
| other | 2 | ccaata(gtga>deleted)cgtcat | 3bp > deletion > 12 bp |
| | | tctgta(ccgggtgggga>deleted)ggtctg | 3bp > deletion > 12 bp |
| *sup-10* | | | |
| deletions > 5kb | 11 | | |
| substitutions | 3 | cagttc(t>a)cttgta; L>H | |
| | | tggaat(a>g)tggtcgg; M>V* | |
| | | agccag(g>t)tttgta;; splice site mutation | |
| unknown | 2 | | |

*also tctttt(t>c)caacca in intron 150 bp upstream

**Table S5. Sequence analysis of reversion mutants for *polq-1; unc-93(e1500).***

| *polq-1* | | | |
|---|---|---|---|
| *unc-93* | | | |
| deletions > 5kb | 6 | | |
| substitutions | 12 | tgcgga(c>a)aagtcg; Q>K<br>cgttga(c>a)gattttc; T>K<br>gatctc(g>a)gatctg; G>R<br>ttccat(c>t)atttat; S>L<br>tttcta(c>a)ctcatg; T>N<br>tttcat(g>t)attgta; M>I<br>ggggag(c>a)caaatg; A>D<br>aagtcg(t>a)cggaaa; V>D<br>tccttt(c>t)gagaca; R>stop<br>tctata(c>a)attgtc; Y>stop<br>aatata(t>a)ttgctg; Y>stop<br>tgttag(g>a)taatgg; splice site mutation | |
| other | 2 | acgtca(ca>deleted)gttgaa<br>ttttac(t>deleted)ttttag | other<br>microsatellite |
| *sup-9* | | | |
| deletions > 5kb | 20 | | |
| substitutions | 7 | tcttcg(g>a)gctcac; G>E<br>gggtac(c>a)agtgga; Q>K<br>gtggag(c>a)atttta; A>E<br>ccattg(g>a)gactta; G>stop<br>aggcta(c>a)ggtcat; Y>stop<br>tccctg(c>t)aaactc; Q>stop<br>caagta(c>a)aacatg; Y>stop | |
| *sup-10* | | | |
| deletions > 5kb | 55 | | |
| substitutions | 1 | atgtta(a>t)tataag; N>I | |
| other | 1 | gtgatg(a>deleted)catcaa | hairpin |
| unknown | 7 | | |

**a**




**b**

| Genotype | # analyzed plates | Mutants found |
|---|---|---|
| N2 | 340 | 0 |
| *polh-1(ok3317)* | 340 | 0 |
| *polh-1(lf31)* | 340 | 0 |
| *polk-1(lf29)* | 340 | 0 |
| *polh-1(ok3317);polk-1(lf29)* | 740 | *dpy(3); ste(1); let(15); him(6)* |
| *polh-1(lf31);polk-1(lf29)* | 340 | *dpy(3); let(5); him(3)* |
| *msh-6* | 300 | 20 visible mutants |

**Figure S1. Occurrence of spontaneous visible mutants in TLS defective strains. a,** Experimental set-up to determine spontaneous mutagenesis: the F1 brood of non-mutant segregating hermaphrodites (P0) were singled to establish individual populations. These were inspected for mendelian segregation of abnormal phenotypes indicating the occurrence of a recessive mutations in the gametes of the P0. Mutants affecting body morphology (e.g. dumpy/dpy) or movement (i.e. uncoordinated/unc) can be scored in the F2 progeny. Mutations in essential genes (i.e. lethal/let) give rise to islands of dead eggs when populations are allowed to clear the food supply. Elevated numbers of males in the F2 progeny indicate a high incidence of males (him) phenotype, arguing for a dominant him mutation in the F1. **b,** Quantification of visible mutant phenotypes. The data for *msh-6* mutants have been published previously[6].

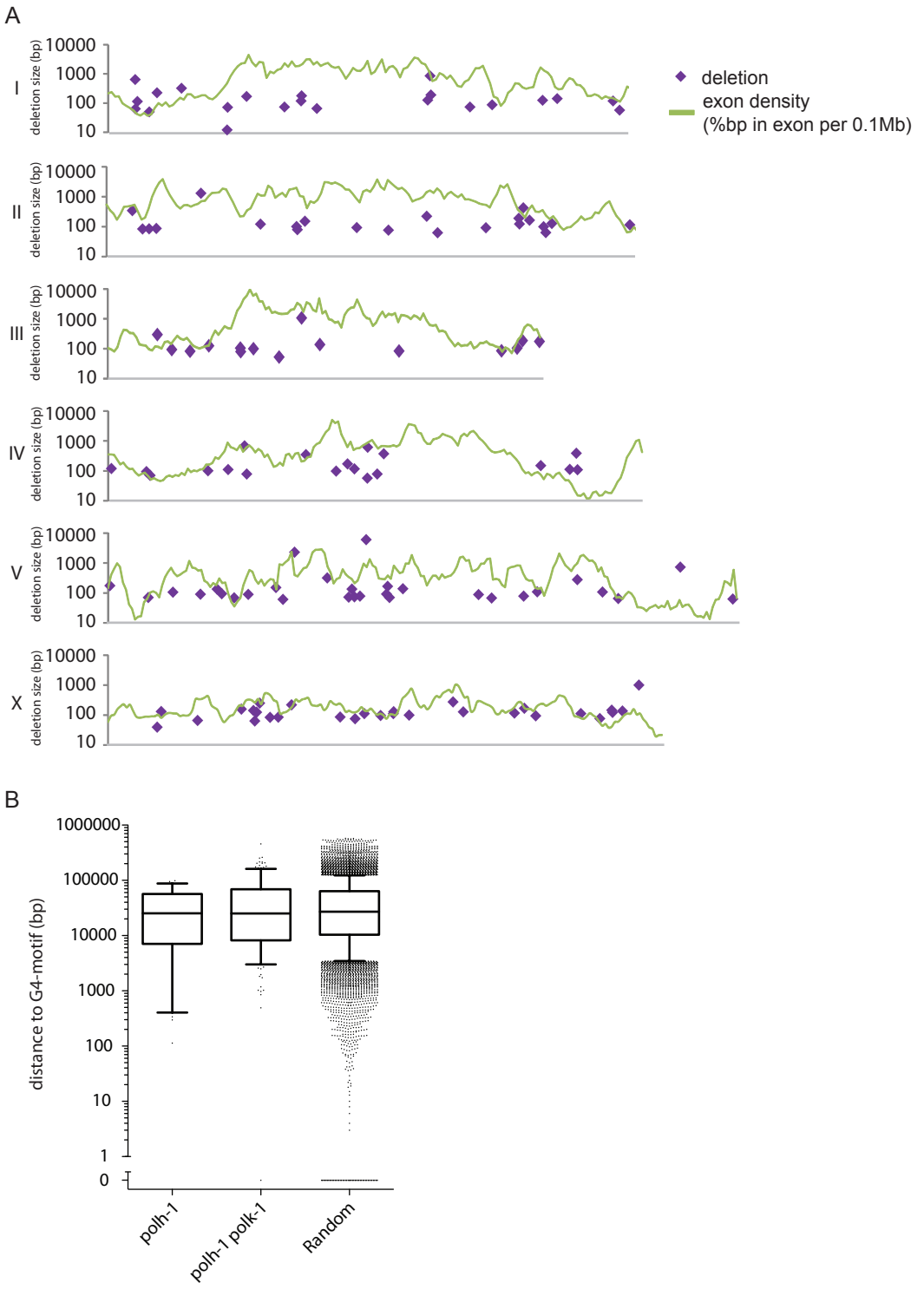



**Figure S2. Genomic distribution of deletions in *polh-1polk-1* mutant animals.**
(A) Individual deletions (purple) were plotted onto a physical map of the *C. elegans* genome. The y-axis shows the size of the deletion on a logarithmic scale. The exon density is displayed in green (y-axis not shown). The length of the graph shows the size of the indicated chromosome relative to each other. (B) For each individual deletion the distance to the closest G4-motif G3-5N1-5G3-5N1-5G3-5N1-5G3-5 (1680 G4-motifs are present in the *C. elegans* genome) was determined. A random set of ~ 13000 deletions with a size distribution similar to those observed in *polh-1 polk-1* mutants was plotted as a comparison. No statistical difference was found between this random set and the set obtained from *polh-1* or *polh-1 polk-1* double mutant animals. Whiskers are drawn down to the 10th percentile and up to the 90th percentile. A distance of zero means that the nearest G4 motif is within the deletion.

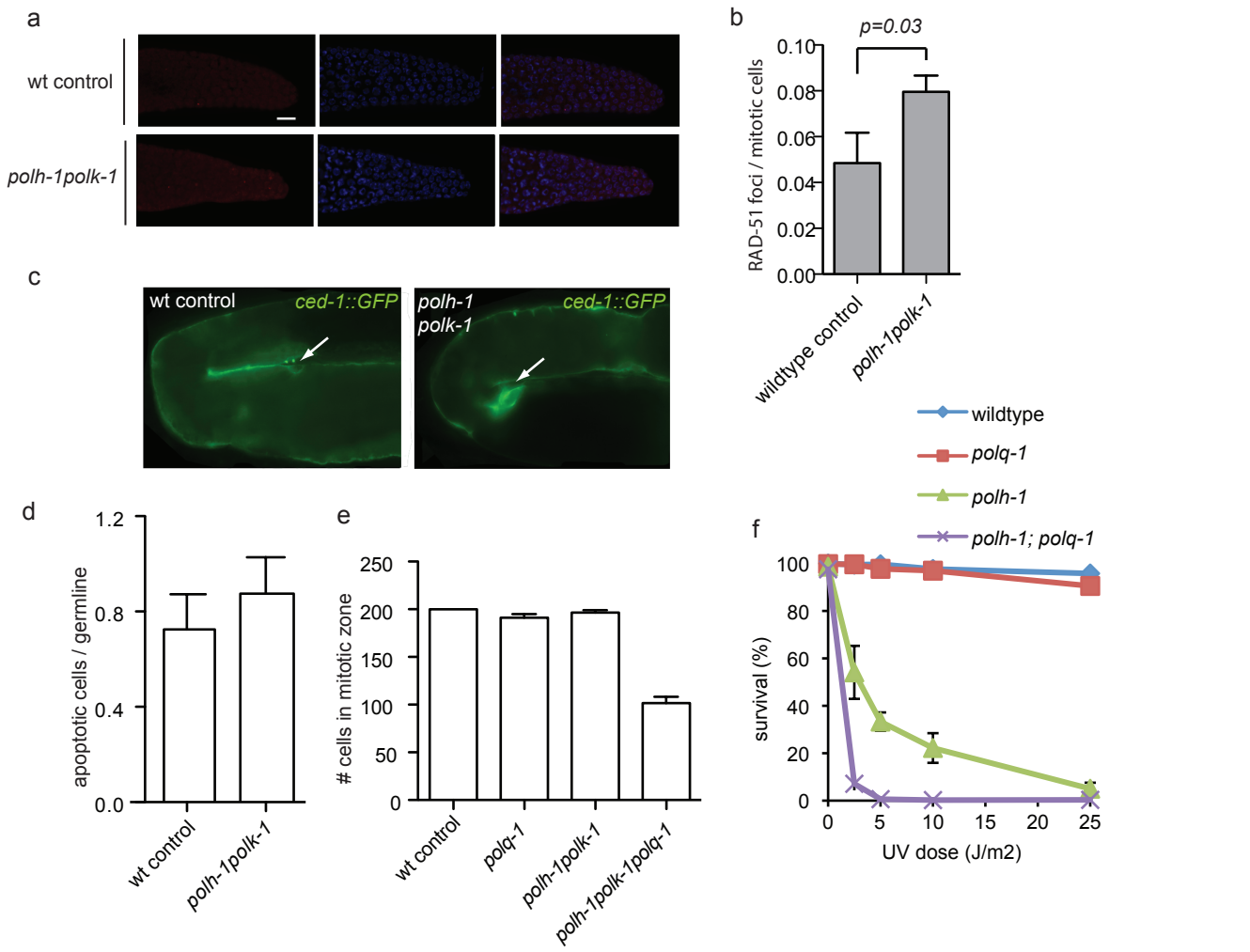



**Figure S3. Analysis of DNA damage induction and apoptosis in single, double and triple mutants of *polh-1*, *polk-1* and *polq-1*. a.** Representative images and **b.** quantification of RAD-51 foci for the indicated genotypes in nuclei present in the proliferative compartment of the *C. elegans* reproductive system. DAPI stainings inblue, RAD-51 in red. Scale bar, 10 µm **c.** Representative images of the bend of the gonad arm of animals transgenic for the apoptotic marker ced1::GFP; cells in the process of apoptotic engulfment are indicated with arrows. Scale bar, 10 µm **d.** Quantification of apoptotic cells in polh-1polk-1 mutant animals and wildtype controls. **e.** Quantification of the number of nuclei in the mitotic region of the germline. A reduction in the number of cells in this region is an established outcome of checkpoint activation. **f.** Sensitivity of *polh-1* and *polq-1* single and double mutants for exposure to UV, plotted as the fraction of surviving progeny after germline exposure of young adult worms.

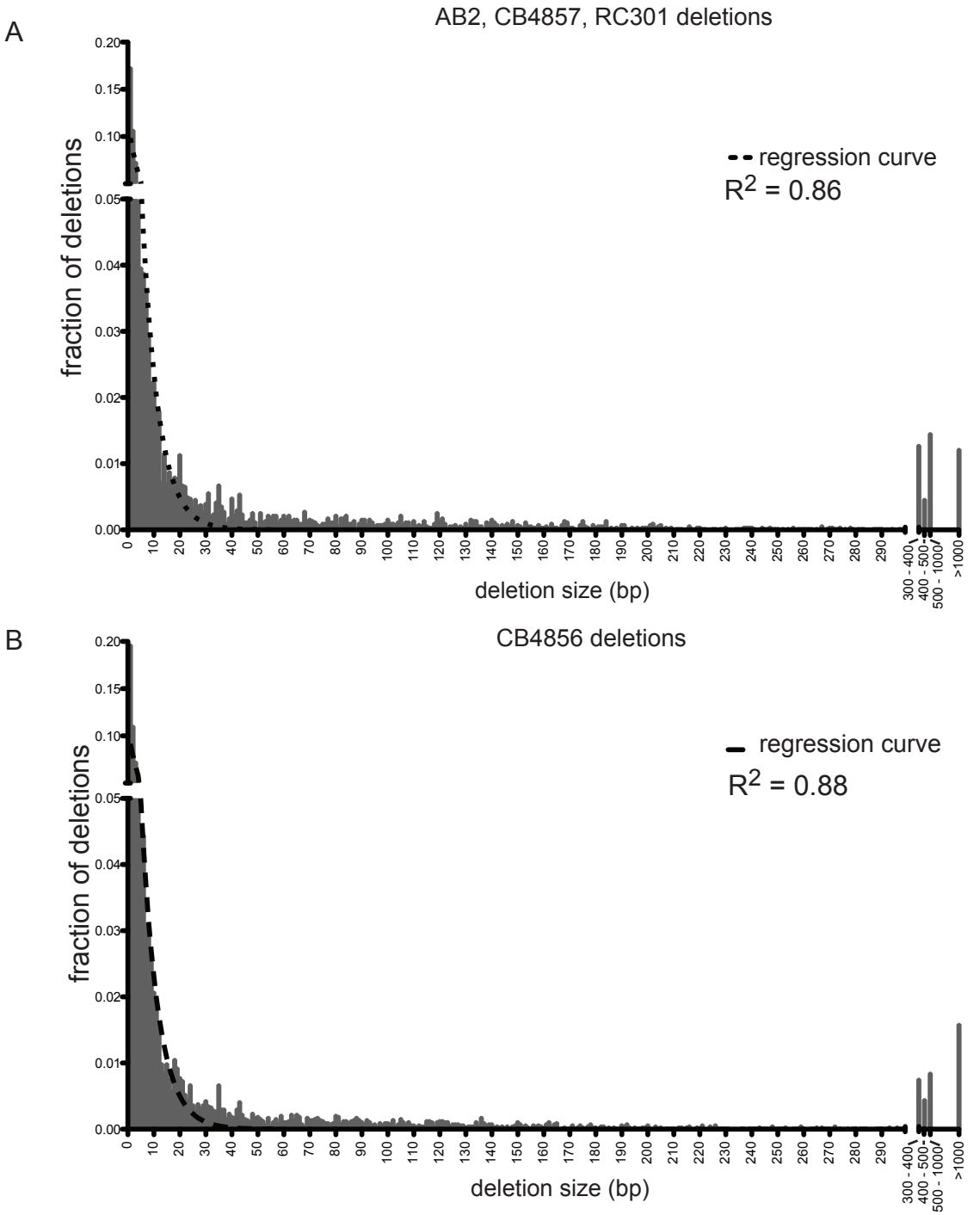
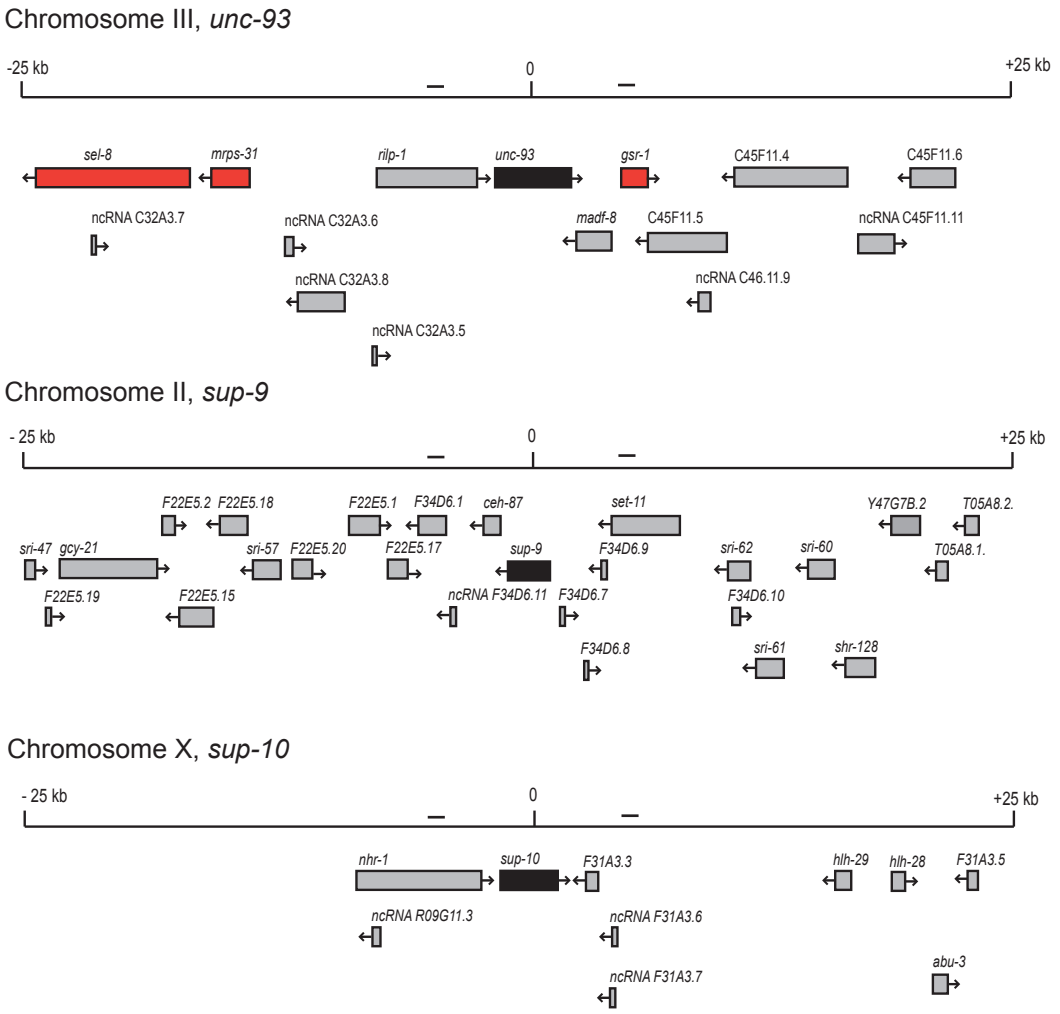



**Figure S4. Histogram of size distributions is plotted of the various *C. elegans* natural isolates that were analyzed.** Regression analysis showed that an exponential fit for deletion sizes up to 20bp approaches the actual distribution best. (**A**) the grouped distribution for AB2, CB4857 and RC301. (**B**) as in (**A**), but now for CB4856.

**A**




**B**

| | *unc-93* | *sup-9* | *sup-10* |
|---|---|---|---|
| wildtype | 0 | 2 | 11 |
| *polq-1* | 6 | 20 | 55 |

**Figure S5.Selective occurence of large chromosomal deletions in regions that are devoid of essential genes in the *unc-93* mutagenesis assay.**(**A**) Schematic representation of 50 kb regions surrounding the *unc-93*, *sup-9* and *sup-10* genes. Known essential genes are depicted in red. While *unc-93* is flanked by two essential genes, no essential genes are known in the 50 kb intervals around *sup-9* and *sup-10*. To estimate deletion sizes, amplification of PCR products at -5kb and +5kb positions has been tested. (**B**) Number of deletions larger than 5kb in *unc-93*, *sup-9* and *sup-10*.